Large-deviation properties of the extended Moran model

Alexander K. Hartmann[1] and Thierry Huillet[2]

[1]Institut für Physik, Universität Oldenburg, D-26111 Oldenburg, Germany
[2]Laboratoire de Physique Théorique et Modélisation, Université de Cergy-Pontoise, 2 Avenue Adophe Chauvin, 95302 Cergy-Pontoise

The distributions of the *times to the first common ancestor* t_{mrca} is numerically studied for an ecological population model, the *extended Moran model*. This model has a fixed population size N. The number of descendants is drawn from a beta distribution $\mathrm{Beta}(\alpha, 2 - \alpha)$ for various choices of α. This includes also the classical Moran model ($\alpha \to 0$) as well as the uniform distribution ($\alpha = 1$). Using a statistical mechanics-based large-deviation approach, the distributions can be studied over an extended range of the support, down to probabilities like 10^{-70}, which allowed us to study the change of the tails of the distribution when varying the value of $\alpha \in [0, 2]$. We find exponential distributions $p(t_{\mathrm{mrca}}) \sim \delta^{t_{\mathrm{mrca}}}$ in all cases, with systematically varying values for the base δ. For $\alpha = 0$ and $\alpha = 1$ the analytically know results $\delta = \exp(-2/N^2)$ and $\delta = 2/3$ are recovered, respectively.

PACS numbers:

I. INTRODUCTION

Population models [1] have attracted the attention in science since many decades [2]. As for any probabilistic model, one is interested in the behavior of random variables defined through the model. In the present work the quantity of interest is the *time to the most recent common ancestor*, which describes how fast a population evolves. Often, in exact calculations as well as in numerical studies, one is restricted to studying averages (or fluctuations) of these quantities. Nevertheless, to obtain a full description of a model, one would like to obtain the full distribution. When performing numerical studies, sophisticated so called *large-deviation approaches* [3–9] can be used to obtain, with some additional numerical effort, probability distributions over a large fraction or even the full range of the support. Here we apply a statistical mechanics-based large-deviation approach to obtain the distribution of the time to the most recent common ancestor for a certain class of population models. This allowed us to calculate the distribution in a region where the probabilities are as small as 10^{-70}.

A particular simple class of models was introduced by Cannings [10, 11]. These population models are simple because they exhibit fixed population size $\mathbb{N} \in \{1, 2, ...\}$. These models are characterized by a family of random variables $\nu_i(t)$, $t \in \mathbb{Z} = \{..., -1, 0, 1, ...\}$, $i \in \{1, ..., N\}$, where $\nu_i(t)$ denotes the number of offspring of individual i of generation t. Since we are not interested in the faith of selected individuals, but only in the evolution of the structure of the population, it is assumed that for each generation t the offspring variables $\nu_1(t), ..., \nu_N(t)$ are exchangeable. We consider a particular subclass of the Cannings population models in which in each generation only one of the N individuals, the *super parent*, is allowed to have more than one offspring. More precisely, our model is defined in terms of a family of random variables $U_N(t) \in \{0, ..., N\}$, which denotes the number of offspring of the super parent in generation t. The model is defined as follows. For $t \in \mathbb{Z}$ and $i \in \{1, ..., N\}$ define

$$\mu_i(t) := \begin{cases} 1 & \text{if } i \in \{1, \dots, N - U_N(t)\}, \\ U_N(t) & \text{if } i = N - U_N(t) + 1, \\ 0 & \text{if } i \in \{N - U_N(t) + 2, \dots, N\} \end{cases} \quad . \quad (1)$$

Now let $\nu_1(t), ..., \nu_N(t)$ be a random permutation of $\mu_1(t), ..., \mu_N(t)$. For each fixed $t \in \mathbb{Z}$ the random variables $\nu_1(t), ..., \nu_N(t)$ are then exchangeable and we interpret $\nu_i(t)$ as the number of offspring of individual i of generation t of a corresponding exchangeable Cannings model.

It is assumed that, for each fixed $N \in \mathbb{N}$, the random variables $U_N(t)$, $t \in \mathbb{Z}$, are independent and identically distributed. The most celebrated example is the standard Moran model [12] corresponding to $U_N \equiv 2$, in which the super parent has two offspring, one other randomly selected individual has no offspring and all the other $N - 2$ individuals have exactly one offspring.

Here, we consider an extended Moran model [13] for the case where $U_N = Nr$ and the random number $r \in [0, 1]$ is drawn from a beta-distribution $\mathrm{Beta}(\alpha, 2 - \alpha)$. The "reproduction parameter" α allowed us to interpolate between the case where the super parent has only a small number of offspring ($\alpha \to 0$) and between the case where the super parent takes over the population very quickly ($\alpha \to 2$). Also the case of a uniform distribution is included ($\alpha = 1$).

II. ALGORITHMS

In the first of the following two subsections, we explain how we simulated the extended Moran model in order to measure the time t_{mrca} to the most recent common ancestor. This is pretty straightforward.

To obtain the distribution in ranges where the probabilities are as small as 10^{-70}, we used a large-deviation algorithm presented below in the second subsection. It consists of a separate level of a Markov Chain Monte Carlo (MCMC) simulation, on top of the simulation of the Moran model.

We present it in a general way such that it can be applied to the simulation an arbitrary "target" stochastic process. Each

instance of a simulation of the target process is assumed to result in a measurable scalar quantity W of interest, exhibiting a probability distribution $P(W)$ (below W will be the time t_{mrca} to the most recent common ancestor). On a digital computer, stochastic processes can be simulated using (pseudo) random numbers, denoted here as $\{\xi_i\}$. Usually, the (pseudo) random number are computed on the fly while the target simulation is performed. Equivalently, one can *precompute* (or obtain in a different way) the random numbers before the actual simulation is performed [14]. This *set* of random numbers is stored in a vector $\xi = (\xi_0, \xi_1, \ldots, \xi_{M-1})$ of suitable length M. While the actual target process is simulated, the random numbers used are taken from the set ξ. Therefore, ignoring the dependence on initial conditions etc, the outcome W of the target process depends only *deterministically* on the set of used randoms numbers, i.e., $W = W(\xi)$. Most general, the entries of ξ are random variables uniformly distributed in $[0, 1]$, since any type of random numbers can be obtained from them via suitable transformations.

Note that, to ensure a good convergence behavior of the algorithm, one should use each entry of ξ always for the same purpose, independently of other entries of ξ. This means, as we will see below, some values of ξ will be ignored sometimes. This helps to ensure that a small change in ξ leads typically to a small change of W, which is necessary for a good behavior of the algorithms used below.

A. Simulating the extended Moran model

As introduced above, we assume that the random numbers, uniformly $U(0, 1)$ distributed, needed for performing one run (population size N and $t_{\max}$ generations) are stored in a vector $\xi = (\xi_0, \xi_1, \ldots, \xi_{M-1}) \in [0, 1)^M$ of size $M = t_{\max} \times (N + 2)$. Thus, to simulate a single generation (up to) $N + 2$ random numbers are needed.

Here, the evolution starts with the population at time $t = 0$. The evolving fate of the population is stored as vectors $a(t) = (a_1(t), \ldots, a_N(t))$ of numbers denoting the corresponding ancestors at the initial time $t = 0$. Thus, the value $a_i(t)$ denotes in generation t the ancestor of individual i from generation 0. Therefore, the vector is initialized as $a_i(0) = i$ for $i = \{1, \ldots, N\}$.

The evolution of the population is simulated forward in time. In each generation $t = 1, \ldots, t_{\max}$, one super parent $n_0 \in \{1, \ldots, N\}$ is selected randomly and uniformly in the population. For this purpose, the entry $\xi_{t(N+2)}$ is used. Thus, the random selection is achieved via simply setting $n_0 = \lfloor N\xi_{t(N+2)} \rfloor$. Next, the number $U_N(t)$ of offspring of n_0 is selected via drawing a random number from the Beta$(\alpha, 2 - \alpha)$ distribution. Drawing this random number works using the inversion method [15], for this purpose the entry $\xi_{t(N+2)+1}$ is used. The generated number is multiplied by $N + 1$ and the floor is taken, resulting in $U_N(t) \in \{0, \ldots, N\}$. Note that if $U_N(t) = 0$ we define that still individual n_0 will generate one offspring by the next step: From the population at time $t - 1$ those $N - U_N(t)$ members n have to be selected (uniformly), which have exactly one offspring, i.e., $a_n(t) = a_n(t-1)$. For

this purpose the entries $\xi_{t(N+2)+2}, \ldots, \xi_{t(N+2)+1+N-U_N(t)}$ are used, i.e., none if $U_N(t) = N$. Thus, the subsequent entries $\xi_{t(N+2)+1+N-U_N(t)}, \ldots, \xi_{(t+1)(N+2)-1}$ are not used for the simulation. Finally, the members n of the population, which were *not* among those selected for having exactly one offspring, are selected as offspring of individual n_0, i.e., $a_n(t) = a_{n_0}(t - 1)$. By this, the calculation of generation t is complete.

t	a(t)								n_0	U_N
0	1	2	3	4	5	6	7	8	4	3
1	1	4	3	4	4	6	7	8	6	2
2	1	4	3	4	6	6	7	8	2	4
3	1	4	3	4	6	4	7	4	6	5
4	4	4	4	4	4	4	4	4		

FIG. 1: Sample evolution for a population with $N = 8$ individuals. Shown are the number t of generations (left), the values of the ancestors $a_1(t), \ldots, a_N(t)$, the selected super parent n_0 for the next generation, and the number $U_N(t)$ of offspring of the super parent. In this case, after four generations, all individuals of the populations are descendants of individual 4 of the initial generation.

In Fig. 1 one sample evolution is shown. As visualized in the figure, at some random time, for a finite population size N and if the total simulation time $t_{\max}$ is large enough, all members population are for the first time offspring of the same individual which was present in the initial population ($t = 0$). Thus, they all have a common ancestor, i.e., $a_n(t_{\min}) = a_1(t_{\min})$ for $n = 1, \ldots, N$. We say the individual *dominates* the population. Thus, looking from this time backward to the initial configuration, which is statistically equivalent to looking forward, this time is the time t_{mrca} to the most recent common ancestor we are interested in, i.e., $t_{\mathrm{mrca}} = t_{\min}$

Note that we measure the probability $P(t)$ that *any* individual dominates (for the first time) after $t = t_{\min}$ steps, which means we look forward in time. This probability is interestingly the same as the probability $P(i \, \mathrm{dom.} \, \mathrm{after} \, t | i \, \mathrm{dom.})$ that a *specific* (say $i = 1$) individual dominates after t steps, conditioned that it is individual i which dominates after some time, which basically means one starts at the time where one specific individual dominates, and looks backwards in time. This can be seen easily, because:

$$P(t) = \sum_i P(i \, \mathrm{dom.} \, \mathrm{after} \, t | i \, \mathrm{dom.}) P(i \, \mathrm{dom.}) \quad (2)$$

$$= N P(1 \, \mathrm{dom.} \, \mathrm{after} \, t | 1 \, \mathrm{dom.}) \frac{1}{N} \quad (3)$$

$$= P(1 \, \mathrm{dom.} \, \mathrm{after} \, t | 1 \, \mathrm{dom.}) \quad (4)$$

because by symmetry of all individuals $P(i \, \mathrm{dom.}) = \frac{1}{N}$ and

the conditional probabilities are all the same, thus the sum reduces to a multiplication with N.

One can measure, for example, the mean of t_{mrca} and obtain a (small-support) histogram via *simple sampling*: One generates, say, K times a vector ξ of random numbers, runs each time the above described algorithm to generate a dynamical of the evolution. Then one measures for each run the resulting time where for the first time all members of the population originate from the same ancestor. Depending on the value K of independent runs, the obtained histogram will be of better or worse quality. Typically, Probabilities of the order of $1/K$ can be measured, like 10^{-9}.

B. Large-deviation approach

Following the description so far, one is able to simulate the evolution of the population in a standard way, only the generation of the random numbers and the actual simulation of the stochastic target process are disentangled.
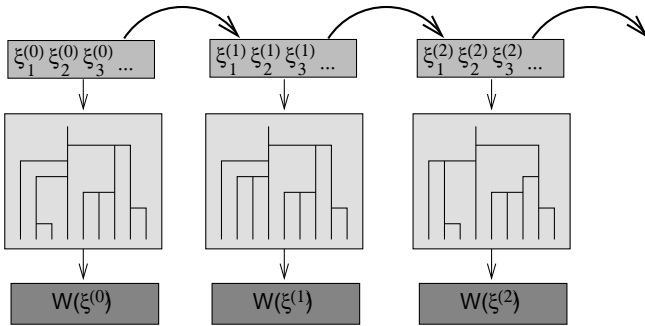

FIG. 2: Schematic representation of the evolution of the set $\xi^{(s)}$ (s being the Monta Carlo step), where each set $\xi^{(s)}$ feeds a target process (middle row, here a simulation of instance of the extended Moran model) resulting in a measurable quantity W (here the time to the most recent common ancestor). Thus, W depends in this way deterministically only on the current set ξ, i.e., $W = W(\xi)$, while the evolution of $\xi = \xi^{(s)}$ is governed by a stochastic process.

Nevertheless, this disentanglement allows one to perform a Markov-chain Monte Carlo evolution for the set ξ: $\xi^{(0)} \rightarrow \xi^{(1)} \rightarrow \xi^2 \rightarrow \ldots$. Thus, the target process is not performed several times independently, but for a sequence of similar (correlated) sets $\xi^{(s)}$ ($s = 0, 1, 2, \ldots$), see Fig. 2. This may appear inefficient on the first sight since the measured quantities $W(\xi^{(0)})$, $W(\xi^{(1)})$, $W(\xi^{(2)})$, ... will be correlated as well, in contrast to using each time a new set ξ, corresponding to *simple sampling*. On the other hand, the MC evolution allows one to introduce a bias, such that the measured distribution for W can be directed into a region of interest, e.g., where the original probabilities $P(W)$ are very small. Here, since we are interested in the distribution of W for a large range of the support, an exponential Boltzmann bias $e^{-W/\Theta}$ is used where Θ is a freely adjustable parameter (a kind of artificial temperature), which allows us to center the observed distribution in different regions. Note that the choice $\Theta = \infty$ corresponds to

```
algorithm large deviation distributions
begin
    Generate initial set vector ξ, each entry uniform in [0, 1]
    Perform target process using set vector ξ
    measure quantity W, i.e. W = W(ξ)
    for trial = 1, . . . , n_{MC−trial}
    begin
        ξ' = change some entries of ξ randomly
        measure quantity W(ξ')
        if W(ξ') "valid": accept ξ' with Metropolis probability
            min{1, exp(−(W(ξ') − W(ξ))/Θ)}
    end
end
```

FIG. 3: Summary of the large-deviation MC algorithm (see text)

the absence of the bias, i.e., to the simple sampling presented in Sec. II A, only including correlations.

The bias is included in the following way via using the Metropolis-Hastings algorithm. Each step s consist of changing a small number of randomly selected entries of the set $\xi^{(s-1)}$, the value of each drawn again uniformly from $[0, 1]$. This results in a "trial set" ξ' with corresponding quantity $W(\xi')$. For some applications, $W(\xi')$ may not be "valid" (e.g., when the time t_{mcra} exceeds the maximum number $t_{\max}$ of generations). In this case, ξ' is immediately *rejected*, which means $\xi^{(s)} = \xi^{(s-1)}$. Otherwise, the trial set ξ' is *accepted* ($\xi^{(s)} = \xi'$) with the Metropolis probability $\min\{1, \exp(-(W(\xi') - W(\xi^{(s)}))/\Theta)\}$. If not accepted, ξ' is also rejected. The algorithm is summarized in Fig. 3.

The performance of the algorithm is influenced by the number of entries of the set ξ which are changed to generate the trial set ξ'. As a good rule of thump, this number should be chosen such that the empirical acceptance rate of the trial configuration is around 0.5.

Since the generation of the random numbers is disentangled from the target process, this approach does not require the introduction of a "reaction coordinate", which is a quantity used to guide certain types of large-deviation simulations. This is in contrast to many standard large-deviation algorithms for dynamical equilibrium and non-equilibrium systems [16], like transition-path sampling [17]. Thus, within the present approach the target process leading to the measurable quantity W can be seen as a black box. This allows to perform large-deviation measurements for almost arbitrary equilibrium and non-equilibrium processes which can be simulated on a computer using (pseudo) random numbers.

In Fig. 4 two sample evolutions of t_{mrca} are shown as a function of the number s of Monte Carlo steps. Both evolutions are for a population size $N = 200$, reproduction parameter $\alpha = 1$ and MC temperature $\Theta = -2.5$. The negative temperature results in value of t_{mrca} which are larger than typical values for this choice of N and α. Note that the two data sets start from opposite side: one set is for a initial configuration vector $\xi^{(0)}$ which is drawn independently in $[0, 1)^M$. This corresponds to the typical behavior. The other vector is drawn such that it exhibits a preference for large values of
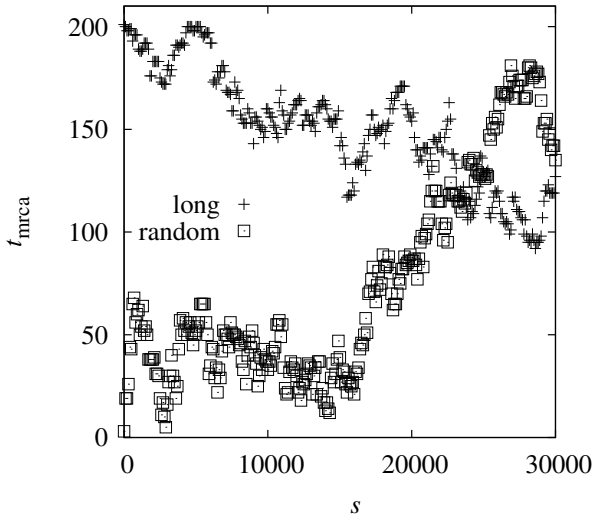
FIG. 4: Time t_{mrca} to most recent common ancestor as function of the Monte Carlo time s for population size $N = 200$, reproduction parameter $\alpha = 1$, sampling temperature $\Theta = -2.5$ for two different initial start configuration $\xi^{(0)}$: one completely random and one with a bias towards larger values of t_{mrca}.

t_{mrca}. This is achieved by sampling the entries which are responsible for drawing the number of offspring not uniformly in the interval $[0, 1]$ but in a smaller interval $[0, 0.15]$. In spite of the different initial conditions, after a while the value of the measured quantity t_{mrca} agree within the range of the fluctuations. This indicates that the Markov chain has "forgotten" its initial condition, i.e., can be considered as *equilibrated*.

For each value of Θ, one obtains a distribution which includes the original distribution under the bias $e^{-W/\Theta}$. Therefore it is related to the original distribution [6] via

$$P(W) = e^{W/\Theta} Z(\Theta) P_{\Theta}(W) \qquad (5)$$

where $Z(\Theta)$ is the normalization constant of $P_{\Theta}(W)$, which can be determined from the numerical data, as explained next. By performing the simulation for suitable chosen values of Θ, which have to be determined experimentally, one can cover a broad range of the desired distribution $P(W)$. If this is done such that the resulting distributions $P_{\Theta}(W)$ overlap for neighboring values of the temperature, say Θ_1 and Θ_2, one can reconstruct the relative normalization constants from $e^{W/\Theta_1} Z(\Theta_1) P_{\Theta_1}(W) = e^{W/\Theta_2} Z(\Theta_2) P_{\Theta_2}(W)$. Actually, several values of W in the windows $[W_{\Theta_2}^{\min}, W_{\Theta_1}^{\max}]$ of overlap are considered and the mean-square difference in this window between the distributions $P(W)$ obtained from (5) is minimized to obtain the "optimum" relative normalization constant $Z(\Theta_1)/Z(\Theta_2)$. This "gluing together" of the distributions of different values of Θ is similar to the approach of Ferrenberg and Swendsen [18] (where the reweighting was not to infinite temperature, like here, but to finite temperatures close to the sampled ones). The last missing constraint is obtained from the overall normalization of $P(W)$ which then results in all normalization constants. Details can be found in Ref. [6], a generalization is explained in Ref. [19].

III. RESULTS

Simulation have performed for different population sizes N, different values of the reproduction parameter α and various temperatures Θ. The values of different temperatures for a given combination N, α ranged from two to seven.

For each combination of these parameters, the number of how many of the entries of the configuration vector ξ were redrawn to obtain the trial vector, was somehow adopted. As a rule of thump (often used for Monte Carlo simulation), this amount of change was chosen such that the acceptance rate of the Metropolis steps was roughly near 0.5. Note that the number c_{sp} of changes for the entries of ξ which are responsible for selecting the super parent, for selecting the number of offspring (c_{o}), and for selected the one-offspring parents (c_1), were tuned separately. The reason is that the amount of change for super parent entries and numbers of offspring entries have a higher influence on the acceptance rate than the amount of change for the one-offspring parent entries.
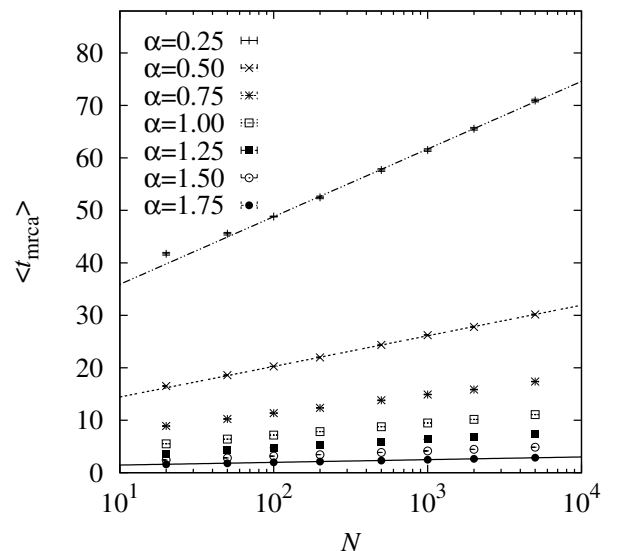

FIG. 5: Mean time t_{mrca} to most recent common ancestor as function of the population size N, for different values of the reproduction parameter α.

Using simple-sampling, we measured the mean of the time t_{mrca} to the most recent common ancestor as a function of the population size N. The result is shown in Fig. 5 using a logarithmically-scaled N-axis. When ignoring very small values of N the data follows straight lines very well, meaning that it is described well by a logarithmic growth. This matches an analytical calculation for the present case [22] very well.

With increasing value of α, i.e., when the distribution of the number of decedents shifts to larger fractions of the population, the mean t_{mrca} decreases in particular the slope. We have fitted the data to function $a + b \log(N)$, for the different values of α. The result is shown in Fig. 6 as a function of $2 - \alpha$. Note that for $\alpha = 1$ we obtained $b = 1.002(4)$ which is compatible with $b = 1$.
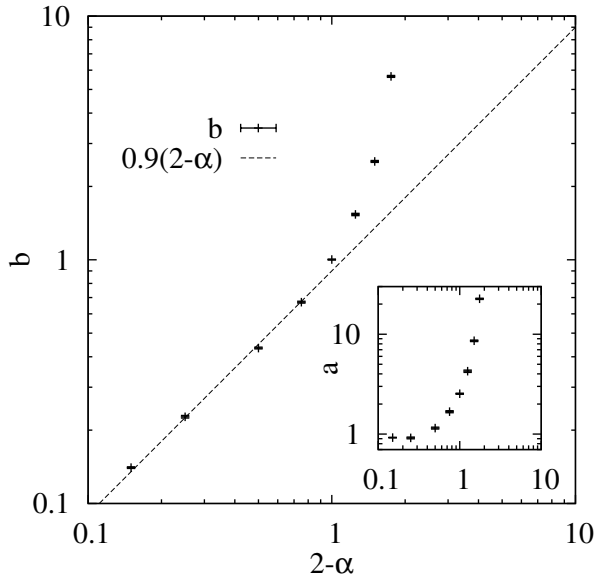
FIG. 6: Fit parameter b when fitting $a+b\log(N)$ N: population size) to the mean time t_{mrca} to most recent common ancestor as function of the population size N, as a function of $2-\alpha$ (with α being the reproduction parameter). The inset shows the behavior of a as function of $2-\alpha$:
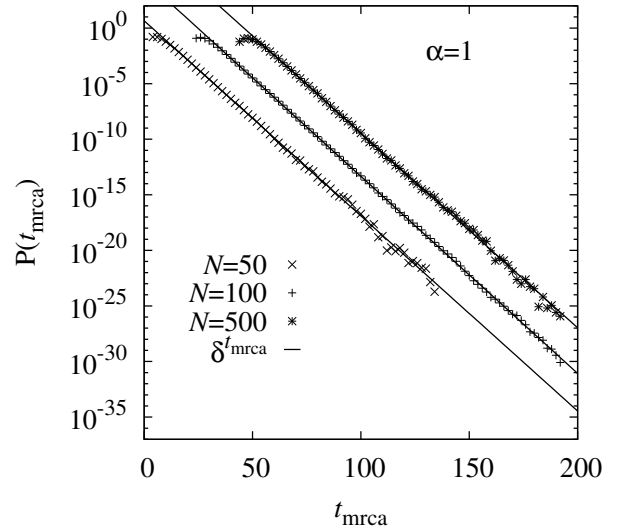

FIG. 7: Distribution of the time t_{mrca} to the most recent common ancestor for reproduction parameter $\alpha = 1$ and three different population sizes $N = 50$, $N = 100$ and $N = 500$. For better visibility, the data for $N = 100$ is shifted by 20 time units to the right, and the data for $N = 500$ by 40 time units. The lines show the results of fitting the tails of the distributions to exponential functions, respectively.
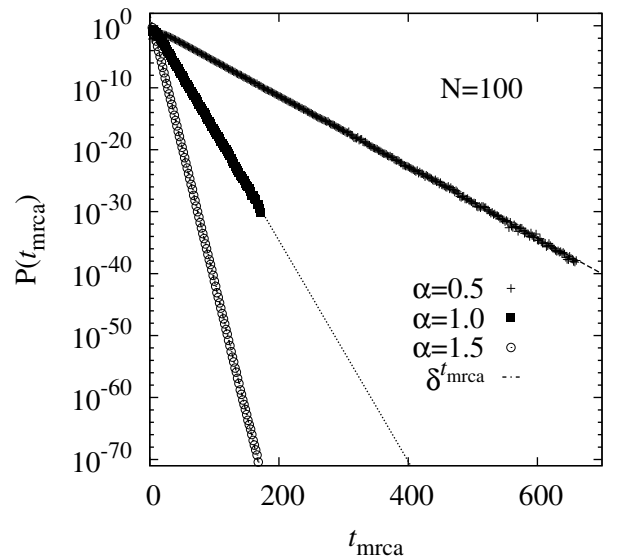
In the log-log plot, a straight line is visible for $\alpha \to 2$, meaning that b converges to zero. Comparably, a appears to converge to a finite value near 1 (see inset of Fig. 6). This is reasonable because for $\alpha \to 2$ the distribution of the number of descendants converges to the case where the super parent takes over the full population in one generation, independent of the size of the population, which means $t_{\mathrm{mcra}} = 1$ with $a = 1$ and $b = 0$.

Next, we study the full distribution of t_{mrca}, see Fig. 7 for the case $\alpha = 1$, i.e., the uniform distribution, and different population sizes N. Using the large-deviation approach, we could measure the distribution over a large range of the support. Apparently these tails exhibit an exponential shape. We fitted these tails to functions $\sim \delta^{t_{\mathrm{mrca}}}$ for all values of N. We obtained $\delta = 0.668(2)$ ($N = 50$), $\delta = 0.665(1)$ ($N = 100$), $\delta = 0.668(1)$ ($N = 500$) (The error bars are just statistical error bars). All these values are very close to the exact value [23] $\delta(1) = 2/3$ and are within the fluctuations not depending on the population size. Thus, even small population sizes are suitable for obtaining results close to the $N \to \infty$ limit, if the tails of the distribution are accessible. For this reason, we proceed with results for $N = 100$, for various values of $\alpha \in [0, 2]$.

In Fig. 8 the distributions for three different values of α and a population size of $N = 100$ are shown. Again, the tails of the distributions can be well fitted to exponential functions $\sim \delta^{t_{\mathrm{mrca}}}$ for all values of α. We did this for all values of α which we have studied.

FIG. 8: Distribution of the time t_{mrca} to the most recent common ancestor for population size $N = 100$ and two example values of the reproduction parameter α.

In Fig. 9, the resulting behavior of δ as a function of α is shown. With increasing value of α, i.e., when the distribution of the number of offspring of the super parent is more and more located at large values, the value of δ decreases, corresponding to smaller times it takes for one individual to dominate the full distribution. Thus, for $\alpha \to 2$ we obtain $\delta \to 0$ corresponding to an $t_{\mathrm{mrca}} = 1$.

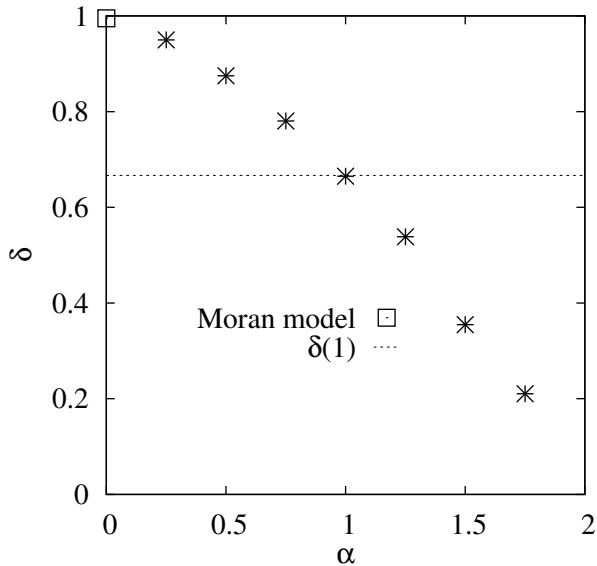FIG. 9: Value of the base δ of the tail $P(t_{\mathrm{mrca}}) \sim \delta^{t_{\mathrm{mrca}}}$ of the distribution of the time to the most recent common ancestor as a function of the reproduction parameter α. The value for the standard Moran model (corresponding to $\alpha \to 0$) is $\exp(-2/N^2)$, which is shown as square symbol. The horizontal line indicates the exact value $\delta = 2/3$ for $\alpha = 1$.

For the opposite limit $\alpha \to 0$ the model converges to the classical Moran model [12]. Here, the distribution of $2t_{\mathrm{mrca}}/N^2$ follows (see Eq. (53) and below in Ref. [20]) in the tails a simple exponential distribution $\exp(-t)$. This corresponds to a value $\delta = \exp(-2/N^2)$ which evaluates for the present case ($N = 100$) to $\delta = 0.995$, i.e. very close to 1, as shown in the figure.

IV. SUMMARY AND DISCUSSION

We have studied the extended Moran model for a beta($\alpha, 2 - \alpha$)-distribution for the fraction of the population which descends from the super parent. In particular we have studied the time t_{mrca} to the most-recent common ancestor. The typical behavior (like the mean) of this quantity can be readily studied using standard numerical simulations.

In order to describe the statistics of this model to a large extend, we investigated not only the typical behavior but also the distribution of t_{mrca} towards the tails. To access these tails numerically, we had to use a statistical mechanics-based biased sampling approach, which is based on a Markov-chain evolution of a vector of uniformly distributed random numbers from the interval $[0, 1]$, respectively, seen as an input vector to an arbitrary stochastic process.

We found that the mean time t_{mrca} depends for all values $\alpha <$ logarithmically on the population size and converges to a constant for $\alpha \to 2$. The distribution of t_{mrca} shows an exponential behavior in the tails. For the cases $\alpha = 1$ and $\alpha = 0$ both mean as well as tail behavior are compatible with previous analytical results.

This work shows that using sophisticated sampling techniques, the distributions of measurable quantities in population models can be studied over large ranges of the support. This allows one to access results in regions where no analytical results are available, as here for the extended Moran model. It would be interesting to apply such techniques to more refined models, such models with varying population size, or models having a spatio-temporal evolution.

[1] M. Kot, *Elements of Mathematical Ecology* (Cambridge University Press, Cambridge, 2008).
[2] O. Aydogmus (2018), arXiv:1801.04597.
[3] F. den Hollander, *Large Deviations* (American Mathematical Society, Providence, 2000).
[4] H. Touchette, Physics Reports **478**, 1 (2009), ISSN 0370-1573.
[5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer, Berlin, 2010).
[6] A. K. Hartmann, Phys. Rev. E **65**, 056102 (2002).
[7] A. K. Hartmann, Eur. Phys. J. B **84**, 627 (2011).
[8] A. K. Hartmann, S. N. Majumdar, and A. Rosso, Phys. Rev. E **88**, 022119 (2013).
[9] A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).
[10] C. Cannings, Adv. Appl. Probab. **6**, 260– (1974).
[11] C. Cannings, Adv. Appl. Probab. **7**, 264– (1975).
[12] P. Moran, Proc. Cambridge Philos. Soc. **54**, 60 (1958).
[13] T. Huillet and M. Möhle, Theor. Popul. Biol. **87**, 5–14 (2013).
[14] G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 (2001).
[15] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual* (Network Theory Ltd., Bristol, UK, 2006).
[16] J. T. Berryman and T. Schilling, J. Chem. Phys. **133**, 244101 (2010).
[17] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[18] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[19] M. R. Shirts and J. D. Chodera, The Journal of Chemical Physics **129**, 124105 (2008).
[20] T. Huillet, Math. Popul. Stud. **23**, 164 (2016).

[21] T. Huillet, Journal of Statistics: Advances in Theory and Applications **7**, 85 (2012).

[22] The mean is stated in general in Ref. [21] after Eq. (37), see page 121. For the present case one has to perform a short calculation by using the corresponding measure which results in a $\log(N)$ dependence.

[23] See bottom of page 145 of Ref.[21].