Lookdown construction for a Moran seed-bank model*

M. C. Fittipaldi[†]   A. González Casanova[‡]   J. E. Nava-Trejo[§]

Abstract

We present a lookdown construction for a Moran seed-bank model with variable active and inactive population sizes and we show that the empirical measure of our model coincides with that of the Seed-Bank-Moran Model with latency of Greven, den Hollander and Oomen [6]. Furthermore, we prove that the time to the most recent common ancestor, starting from N individuals with stationary distribution over its state (active or inactive), has the same asymptotic order as the largest inactivity period. Additionally, we give an explicit approximation of its distribution under extra assumtion on the inactivity and activity switching rates. We then find the first asymptotic order of the fixation time of a single beneficial mutant conditioned to invade the whole population, resulting to be of order $\ln N$.

Keywords: seed bank; lookdown; fixation time; time to the most common ancestor.
MSC2020 subject classifications: Primary 60K35, Secondary 92D15.
Submitted to ECP on April 11, 2023, final version accepted on August 9, 2024.

Introduction and main results

Seed-banks have been studied for their important effects in biology, but they are also of great interest from a mathematical point of view, since they can notably modify certain phenomena. In the probabilistic community, efforts to study the effect of seed-banks rigorously can be traced back to the work of Kaj, Krone and Lascoux [7]. They postulated and studied an extension of the classical Wright-Fisher model which includes a seed-bank. The main difficulty of their approach is that the stochastic processes they study are not Markovian. Overcoming this (in a particular case) was achieved by Blath, Kurt, Wilke-Berenguer and one of the authors of this paper in [1]. There, a new Wright-Fisher model with fully Markovian seed-bank was constructed. Moreover, the authors showed that the genealogy of this new model converges to the Seed-Bank coalescent. This process is a Kingman-type coalescent, where the lines of the coalescent can enter into

latent state. As opposed to the Kingman coalescent, the seed-bank coalescent does not *come down from infinity*. Also in [1] it was proved that the time to the most recent common ancestor of a sample of N active individuals is of order $\ln(\ln(N))$. In [5] the reader can find a detailed description of the Seed-Bank coalescent.

A formidable tool to study coalescent processes is the *lookdown* construction by Donelly and Kurtz [2, 3]. This type of construction allows to build Moran-type models where the lower N levels form a Markov process, for every $N \in \mathbb{N}$. This kind of construction has played a prolific role in the study of many phenomena [4, 9]. Such representations are not known for seed-bank models and one main task of this paper is to fill this gap.

Our starting point is the Moran process with latency of Greven, den Hollander and Oomen [6], in which active and dormant individuals are not two separate families of lines, but instead each line consists on active and dormant periods exponentially distributed with rates α and σ respectively.

In this work, we present a lookdown construction for a Moran model with seed-bank with variable active population sizes and fixed total population size. This is a relaxation of assumption in [1] that the sizes of the active and dormant population are constant: we keep the total population size constant but allow the proportion of active (resp. dormant) individuals to vary over time, starting from a stationary distribution. We show that the empirical measure of our SB-lookdown construction and the empirical measure of the SB-Moran Model presented in [6] coincide via the Markov mapping theorem.

Besides our novel lookdown process, we have two other main results. The first contribution is related to the time to the most recent common ancestor of the seed-bank coalescent. Thanks to [1], it is known that the time to the most recent common ancestor of the seed-bank coalescent starting from an active and bn inactive individuals fulfills

$$\mathbb{E}\left[T_{MRCA}[(an, bn)]\right] \asymp \ln(\ln(an) + bn). \tag{0.1}$$

While this important result is concerned with expectations, we were able to approximate the distribution. Let $p := \frac{\alpha}{\alpha + \sigma}$ be the average time that a line is inactive. Under the condition $\sigma < (1 - p)^2$, we are able to proof that the time to the most recent common ancestor $\widetilde{T}_{MRCA}[n]$ starting from n individuals with stationary state distribution can be approximated by

$$\widetilde{T} \leq \lim_{n \to \infty} \sigma \left(\widetilde{T}_{MRCA}[n] - \frac{1}{\sigma} \ln np \right) \leq \sigma\varrho + \widetilde{T} \qquad a.s.,$$

where $\widetilde{T} \sim Gumbel(0, 1)$ and ϱ is the time to the most recent common ancestor of the Kingman coalescent. To prove this result, we study the seed-bank coalescent as the ancestry process in the lookdown construction. This allowed us to compare $\widetilde{T}_{MRCA}[n]$ with the largest period of inactivity of an individual, denoted by ψ_n, and to prove that they are of the same order. This comparison enables us to improve the existing result (0.1) in [1] by computing explicitly the asymptotic first order of the expected value of $\widetilde{T}_{MRCA}[n]$ given by

$$\lim_{n \to \infty} \frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\frac{1}{\sigma} \ln np} = 1, \qquad \forall \alpha, \sigma > 0.$$

Finally, our last main result consists in finding the first asymptotic order of the fixation time $\widetilde{\tau}^N$ of the type in the lower level for the N-lookdown model. The SB-lookdown construction allows us to compare the fixation time and the TMRCA, to show that

$$\frac{\widetilde{\tau}^N}{\frac{1}{\sigma} \ln Np} \xrightarrow[n \to \infty]{\mathbb{P}} 1 \qquad \text{and} \qquad \lim_{N \to \infty} \frac{\mathbb{E}\left[\widetilde{\tau}^N\right]}{\frac{1}{\sigma} \ln Np} = 1, \qquad \forall \alpha, \sigma > 0,$$

and approximate its distribution in the case $\sigma < (1-p)^2$.

The remainder of the paper is structured as follows. In Section 1 we construct the SB-lookdown and in Theorem 1.2 we show that the empirical measure of the SB-lookdown and SB-Moran Model coincide. In Section 2 we study the SB-coalescent as the ancestry process on the SB-lookdown, and in Theorem 2.5 we find the asymptotic order of the TMRCA. Finally, in Section 3 we deduce the asymptotic order of the fixation time of the type in the first level in Corollary 3.1.

1   Seed-Bank lookdown model

Throughout this work, we will use the framework of multitype Moran Seed-Bank model given by [6, B.3], in which it was shown that the diffusion process as well as the ancestry processes associated with the Moran model coincide with those found in [1]. Let $N \in \mathbb{N}$, the N-**Moran Model with seed-bank**, or N-**particle SB-Moran model**, describes a haploid population of N individuals which evolve as follows. Each individual starts with a *type* in some set $E \subseteq \mathbb{N}$ and a *state* in $S = \{a, d\}$ (*active* and *dormant* respectively), according with an exchangeable distribution. Each active individual becomes inactive at rate α and each dormant individual becomes active at rate σ. Independently, a pair of individuals is uniformly chosen at rate 1. The chosen individuals will interact if both are active. As a product of the interaction, one of them, chosen at random, will reproduce and its descendant will replace the other one. We will denote the N-particle SB-Moran model at time $t \geq 0$ by $W^N(t) := \left(W_1^N(t), \ldots, W_N^N(t)\right)$. Indeed, $W^N(t)$ is a random vector on $(E \times S)^N$.

Remark 1.1. Observe that the dynamics mentioned above are completely symmetric. Therefore, given an initial exchangeable configuration, the SB-Moran model will be exchangeable.

We will now introduce an ordered particle model in the setting of *classical* lookdown constructions [2] whose empirical measure distributions will agree with the SB-Moran model described above. Here, each particle will be attached to a level in $[N] = \{1, \ldots, N\}$. The population evolves as the SB-Moran Model except for the reproduction events, in which the parent will always be the individual with lower level involved in the reproduction event. We will refer to this ordered particle model as the **SB-lookdown model**, denoted at time $t \geq 0$ by $Z^N(t) := \left(Z_1^N(t), \ldots, Z_N^N(t)\right)$, a random vector on $(E \times S)^N$.

In the following, we will formalize the lookdown construction and we will prove the equivalence between the ordered and unordered model via the Markov Mapping theorem [4, 8, 9].

1.1   Seed-Bank lookdown construction

The first step in building the SB-lookdown model is to determine the initial activity and dormancy periods. For each level $i \in \mathbb{N}$ we define the auxiliary process

$$\gamma^i(t) := y_0^i b_0^i + t + \int_0^t \int_0^\infty y \mathcal{N}^i(ds, dy).$$

Here y_0^i is an exponential variable with rate σ, b_0^i is a Bernoulli variable with success probability $\sigma / (\alpha + \sigma)$, and $\mathcal{N}^i(ds, dy)$ is a Poisson random measure with intensity measure $\sigma ds \otimes \alpha e^{-\alpha y} dy$, all of them independent of each other. Let $\tau^i(t) := \inf\{s : \gamma^i(s) > t\}$ be the right-continuous inverse of γ^i which defines the activity and dormancy periods as follows: $a^i := \{t \in \mathbb{R}_0^+ : \tau^i(t) \neq \tau^i(s) \, \forall s \in \mathbb{R}_0^+\}$ and $d^i := \mathbb{R}_0^+ \setminus a^i$. We then define the state process of the level i by $Z_i^{N,S}(t) := a \mathbb{1}_{\{t \in a^i\}} + d \mathbb{1}_{\{t \in d^i\}}$. Note that the state

process for a given level only depends on itself. Moreover, since the process starts from its stationary distribution we have that the probability of being active at any moment is described by its stationary distribution on a

$$\mathbb{P}[Z_i^N(t) = a] = \mathbb{P}[t \in a^i] = \frac{\sigma}{\sigma + \alpha} = 1 - p \qquad \text{for all} \quad t > 0, i \in \{1, \dots, N\}.$$

The second part of this construction arises from considering the interactions among levels. The times in which the individual on level i tries to adopt the current type of an individual on the lower level j are dictated by a Poisson process $\mathcal{C}^{i,j}$ with rate 1, and the processes $\{\mathcal{C}^{i,j}\}_{j<i, i,j \in \mathbb{N}}$ are independent of each other. Hence, the type of the individual at level i at time t is determined by the last possible interaction with individuals at lower levels before that time. Namely, if $(r_n^{ij})_{n \in \mathbb{N}}$ are the occurrence times of the process $\mathcal{C}^{i,j}$ and we set

$$T_t^{ij} := \sup_{n \in \mathbb{N}} \{r_n^{i,j} : r_n^{i,j} \in a^i \cap a^j, r_n^{i,j} \leq t\}$$

as the last interaction time between individuals on level i and j before time t, setting $\sup \emptyset$ as zero by convenience. The type-state of the system at time t is then provided by a function g constructed in the following recursive way. Given a random function $f : \mathbb{N} \to E$ which determines the initial type configuration, we set $g(1, t) := f(1)$. For each level i, given $g(1, t), \dots, g(i - 1, t)$, we define

$$g(i, t) := f(i)\mathbb{1}_{\{T_t^i = 0\}} + \sum_{j < i} g(j, T_t^i)\mathbb{1}_{\{T_t^i = T_t^{ij}\}},$$

where $T_t^i := \sup\{T_t^{ij}, j < i\}$ is the last interaction time that changes the type of the individual on level i. Therefore, the state of the N-particle lookdown Moran seed-bank process at time t is given by the vector $(Z_1^N(t), \dots, Z_N^N(t))$, where

$$Z_i^N(t) := \left(Z_i^{N,E}(t), Z_i^{N,S}(t) \right) = \left(g(i, t), a\mathbb{1}_{\{t \in a^i\}} + d\mathbb{1}_{\{t \in d^i\}} \right),$$

with $S = \{a, d\}$ the state space of each level. The only differences in the dynamic of this process with the SB-Moran seed-bank model construction are found in the reproduction events. In the lookdown model, each individual chooses his parent among the individuals in the lower levels, meanwhile in the original SB model a pair of individuals is chosen to interact, and the parent is selected uniformly.

1.2 Equality in law of the SB-lookdown and SB-Moran models

In this (sub)section we will show that the empirical measure of the SB-lookdown and SB-Moran Model coincide. Define $\mathcal{P}(E \times S)$ as the space of probability measures on $E \times S$ and $D_{\mathcal{P}(E \times S)}[0, \infty)$ the space of càdlàg $\mathcal{P}(E \times S)$-valued functions with the Skorohod topology.

Theorem 1.2. *The laws of empirical measures associated with the SB-Moran model and the SB-lookdown model coincide on $D_{\mathcal{P}(E \times S)}[0, \infty)$.*

The classic way to prove law equalities for lookdown constructions involves an explicit coupling of the ordered and unordered models (see [3, Section 2]). Instead, we will utilize the Markov Mapping Theorem (see [9, Theorem A.15]), which addresses the martingale problem through its generator. To this end, let's introduce the generators of the empirical measures of the processes. Let $\mathcal{B}\left((E \times S)^N\right)$ be the space of bounded measurable functions on $(E \times S)^N$. Given a configuration $z \in (E \times S)^N$ and $f \in \mathcal{B}\left((E \times S)^N\right)$, we

write

$$A_M^N f(z) := \sigma \sum_{i=1}^{N} \left(f\left(\phi_i^d(z) \right) - f(z) \right) + \alpha \sum_{i=1}^{N} \left(f\left(\phi_i^a(z) \right) - f(z) \right) + \frac{1}{2} \sum_{i \neq j} \left(f(\phi_{ij}(z)) - f(z) \right)$$

(1.1)

and

$$A_{LD}^N f(z) := \sigma \sum_{i=1}^{N} \left(f\left(\phi_i^d(z) \right) - f(z) \right) + \alpha \sum_{i=1}^{N} \left(f\left(\phi_i^a(z) \right) - f(z) \right) + \sum_{i<j} \left(f(\phi_{ij}(z)) - f(z) \right)$$

(1.2)

which correspond to the infinitesimal generator of the unordered and ordered models respectively. Here, $\phi_{i,j}(z)$ corresponds to replace z_j by z_i only if $z_i^S = z_j^S = a$; $\phi_i^d(z)$ corresponds to replacing z_i^S by d; and $\phi_i^a(z)$ corresponds to replacing z_i^S by a. Let $z_N \in \mathcal{P}(E \times S)$ be the empirical measure associated to z, given by $z_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i}$. Let π a permutation of $[N]$ we denote z_π as the rearrangement of z according with π. For any $f \in \mathcal{B}\left((E \times S)^N \right)$ we define

$$\widehat{f}(z_N) := \frac{1}{N!} \sum_\pi f(z_\pi),$$

(1.3)

which correspond to the uniform average out of all the permutations of a given configuration such that its empirical measure is z_N.

1.2.1 Proof of Theorem 1.2.

Since

$$|A_M^N f(z)| \leq (2\sigma + 2\alpha + N) N \|f\| \quad \text{and} \quad |A_{LD}^N f(z)| \leq (2\sigma + 2\alpha + N) N \|f\|$$

for all $z \in (E \times S)^N$, with $\|f\| = \sup_{z \in (E \times S)^N} |f(z)|$, to apply the Markov mapping theorem it is enough to prove that the infinitesimal generators defined in equation (1.1) and equation (1.2) coincide under the average operator given by (1.3). To this end, note that the terms correspondent to activation and dormancy events coincide. First, for the deactivation mechanism we have that

$$\frac{1}{N!} \sum_\pi \sigma \sum_{i=1}^{N} \left(f\left(\phi_i^d(z_\pi) \right) - f(z_\pi) \right) = \frac{1}{N!} \sum_\pi \sigma \sum_{i=1}^{N} \left(f\left(\phi_{\pi_i^{-1}}^d(z)_\pi \right) - f(z_\pi) \right)$$

$$= \frac{1}{N!} \sum_\pi \sigma \sum_{i=1}^{N} \left(f\left(\phi_i^d(z)_\pi \right) - f(z_\pi) \right) = \sigma \sum_{i=1}^{N} \left(\frac{1}{N!} \sum_\pi f\left(\phi_i^d(z)_\pi \right) - \frac{1}{N!} \sum_\pi f(z_\pi) \right)$$

$$= \sigma \sum_{i=1}^{N} \left(\widehat{f}\left(z_N + N^{-1}(\delta_{\phi_i^d(z)_i} - \delta_{z_i}) \right) - \widehat{f}(z_N) \right),$$

and similarly the activation mechanism has the form

$$\frac{1}{N!} \sum_\pi \alpha \sum_{i=1}^{N} \left(f\left(\phi_i^a(z_\pi) \right) - f(z_\pi) \right) = \alpha \sum_{i=1}^{N} \left(\widehat{f}\left(z_N + N^{-1}(\delta_{\phi_i^a(z)_i} - \delta_{z_i}) \right) - \widehat{f}(z_N) \right).$$

Finally, for the reproduction mechanism we have that

$$
\frac{1}{N!} \sum_\pi \frac{1}{2} \sum_{i \neq j} \left(f(\phi_{ij}(z_\pi)) - f(z_\pi) \right) = \frac{1}{2} \sum_{i \neq j} \left(\widehat{f}(\mathbf{z}_N + N^{-1} \left(\delta_{z_i} - \delta_{z_j} \right)) - \widehat{f}(\mathbf{z}_N) \right)
$$

$$
= \frac{1}{2} \sum_{i<j} \left(\widehat{f}(\mathbf{z}_N + N^{-1} \left(\delta_{z_i} - \delta_{z_j} \right)) - \widehat{f}(\mathbf{z}_N) \right) + \frac{1}{2} \sum_{i>j} \left(\widehat{f}(\mathbf{z}_N + N^{-1} \left(\delta_{z_i} - \delta_{z_j} \right)) - \widehat{f}(\mathbf{z}_N) \right)
$$

$$
= \sum_{i<j} \left(\widehat{f}(\mathbf{z}_N + N^{-1} \left(\delta_{z_i} - \delta_{z_j} \right)) - \widehat{f}(\mathbf{z}_N) \right),
$$

and we conclude that

$$
A^N \widehat{f}(\mathbf{z}_N) := \sigma \sum_{i=1}^N \left(\widehat{f} \left(\mathbf{z}_N + N^{-1}(\delta_{\phi_i^d(z)_i} - \delta_{z_i}) \right) - \widehat{f}(\mathbf{z}_N) \right)
$$

$$
+ \alpha \sum_{i=1}^N \left(\widehat{f} \left(\mathbf{z}_N + N^{-1}(\delta_{\phi_i^a(z)_i} - \delta_{z_i}) \right) - \widehat{f}(\mathbf{z}_N) \right)
$$

$$
+ \frac{1}{2} \sum_{i \neq j} \left(\widehat{f}(\mathbf{z}_N + N^{-1} \left(\delta_{z_i} - \delta_{z_j} \right)) - \widehat{f}(\mathbf{z}_N) \right)
$$

$$
= \widehat{A_{LD}^N f}(z) = \widehat{A_M^N f}(z).
$$

2 Time to the most recent common ancestor

2.1 SB-coalescent process as the ancestry process of a SB-lookdown particle system

Let's recall the notion of Seed-bank coalescent from [1]. For $k \geq 1$ define $\mathcal{P}_k$ as the set of partitions of $[k]$ and the set of marked partitions of $[k]$ as $\mathcal{P}_k^S = \{(\zeta, \vec{u}) | \zeta \in \mathcal{P}_k, \vec{u} \in S^{|\zeta|}\}$. Let $\pi, \pi' \in \mathcal{P}_k^S$. We denote $\pi \succ \pi'$, if π' can be obtained by merging exactly two blocks carrying the a-flag of π, and the resulting block also carries the a-flag. In a similar way, we denote $\pi \bowtie \pi'$ if π can be constructed by changing the flag of precisely one block of π'.

Definition 2.1 (The Seed-bank k-coalescent, Definition 3.1 [1])**.** *For $k \geq 2$ and $\alpha, \sigma \in (0, \infty)$. The seed-bank k-coalescent $\left(\Pi_t^k \right)_{t \geq 0}$ with seed-bank intensity α and relativity seed-bank size σ is defined to be the continuous time Markov chain with values in $\mathcal{P}_k^S$, with the following transition rates:*

$$
\pi \mapsto \pi' \text{ at rate } \begin{cases} 1, & \text{if } \pi \succ \pi', \\ \alpha, & \text{if } \pi \bowtie \pi' \text{ and one } a\text{-flag is replaced by one } d\text{-flag}, \\ \sigma, & \text{if } \pi \bowtie \pi' \text{ and one } d\text{-flag is replaced by one } a\text{-flag}. \end{cases}
$$

Let $\{(N_t, M_t)\}_{t \geq 0}$ be the block counting process associated to the seed-bank coalescent, that is, (N_t, M_t) correspond to the numbers of a-flag and d-flag blocks on Π_t^k for each $t \geq 0$.

Let's show that the ancestry process of a sample of individuals in the N-SB-lookdown model corresponds to the SB-coalescent. Suppose we sample k individuals at a certain time $T \geq 0$. The initial condition correspond to the singletons of the sampled individuals levels marked according with its state at time T, and we will recover its genealogical information tracing the particle system backwards in time. For each time $t \geq 0$, each block flag will correspond to the state of the lower individual level in the block at time $T - t$. Therefore, it is immediately noticeable that each block with a-flag turns into a d-flag block at rate σ and conversely at rate α. Similarly, coalescence events will

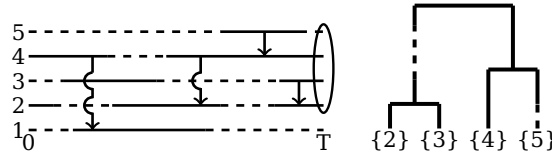correspond to the reproduction events between the lower individuals levels on each block.

Figure 1: An illustration of the 5-particle lookdown model and the ancestry process of a sample of individuals in levels $\{2, \ldots, 5\}$ at time T. Straight lines indicates activity periods meanwhile dashed ones indicates the dormancy periods.

In this context, we present the following result which will be helpful later and it strongly relies on the lookdown construction presented before.

Proposition 2.2 (First collision time with k lower levels.)**.** *Let* $i, k \in \mathbb{N}$*, be such that* $k < i$*. The first reproduction event between the individual at level* i *and any of the individuals at the first* k *lower levels is given by*

$$r_k^i := \inf \left\{ r_n^{ij} \mid r_n^{ij} \in a^i \cap a^j, \, j \in \{1, \ldots, k\}, \, n \in \mathbb{N} \right\}$$

and satisfies $r_k^i \sim \exp\left(k \left(1 - p\right)^2 \right)$*.*

Proof. Since the state process $Z_i^{N,S}$ starts from its stationary distribution recall that $\mathbb{P}[r_n^{ij} \in a^i] = \frac{\sigma}{\alpha + \sigma} = 1 - p$, for all $n \in \mathbb{N}$ and by the independence between levels,

$$\mathbb{P}[r_n^{ij} \in a^i \cap a^j] = \mathbb{P}[r_n^{ij} \in a^i]\mathbb{P}[r_n^{ij} \in a^j] = (1 - p)^2 \qquad (2.1)$$

for $j \in \{1, \ldots, k\}$. Let $\widetilde{C}^{i,k} := \sum_{j=1}^{k} C^{i,j}$ be a Poisson process of rate k by the superposition of Poisson processes. Given K the number of occurrences of $\widetilde{C}^{i,k}$ until the first successful reproduction then we have that $K \sim Geo\left((1 - p)^2\right)$. We can write the distribution of the first successful reproduction event in terms of the number of attempts required for the event to occur, given by K, and the probability of encountering at least K possible reproduction events at time t. This decomposition is expressed as follows:

$$
\begin{aligned}
\mathbb{P}[r_k^i \leq t] &= \sum_{n=1}^{\infty} \mathbb{P}[K = n]\mathbb{P}\left[\widetilde{C}_t^{i,k} \geq n \right] \\
&= \sum_{n=1}^{\infty} (1-p)^2 (1 - (1-p)^2)^{n-1} \int_0^t k \frac{1}{(n-1)!} (ks)^{n-1} e^{-ks} ds \\
&= (1-p)^2 k \int_0^t e^{-ks} \sum_{n=1}^{\infty} \frac{1}{(n-1)!} ((1 - (1-p)^2)ks)^{n-1} ds \\
&= (1-p)^2 k \int_0^t e^{-ks} e^{(1-(1-p)^2)ks} ds = 1 - e^{-k(1-p)^2 t}.
\end{aligned}
$$

$\square$

Remark 2.3. Note that the coalescence between the block that contains i and the one that contains j could be achieved without a reproduction event between their corresponding levels. This implies that if τ_k^i is the coalescent time of the block $\{i\}$ with any of the ancestral lines of the first k individuals, then τ_k^i is bounded from above by r_k^i almost surely.

2.2 Asymptotic distribution of the time to the most recent common ancestor

We will study the long time behavior of the variable

$$\widetilde{T}_{MRCA}[n] = \inf\left\{t \geq 0 \,:\, N_t + M_t = 1 | N_0 + M_0 = n\right\},$$

using the framework developed in [1]. It is important to emphasize that the variable $\widetilde{T}_{MRCA}$ is different from the one defined in [1] as ours it only depends on the total number of sampled initial individuals. The distinction lies in our assumption that the states are in stationarity. In [1, Proposition 4.12], the authors define a stopping time ϱ_n corresponding to *"the first time that all the n initial individuals which so far had not entered the seed-bank have coalesced"*. We can bound the number of lineages left at ϱ_n by the number of lineages that started in a dormant state, plus the number $\mathcal{B}_n$ of first activation periods under the supposition that all individuals started being active, that is

$$N_{\varrho_n} + M_{\varrho_n} \leq \mathcal{B}_{N_0} + M_0 \leq \mathcal{B}_n + M_0,$$

where $\mathcal{B}_k := \sum_{i=2}^{k} \delta_i + 1$, with $\delta_i \sim Be(2\alpha/(2\alpha + (i-1)))$ and are independent of each other, and also independent of M_0. Note that

$$\mathbb{E}[M_0 + \mathcal{B}_n] = pn + \sum_{i=2}^{n} \frac{2\alpha}{2\alpha + (i-1)} = pn + 2\alpha \ln n. + R(\alpha, n), \qquad \text{and}$$

$$\mathbb{V}[M_0 + \mathcal{B}_n] = p(1-p)n + \sum_{i=2}^{n} \frac{2\alpha}{2\alpha + (i-1)}\left(1 - \frac{2\alpha}{2\alpha + (i-1)}\right) \leq E[M_0 + \mathcal{B}_n] \tag{2.2}$$

with $R(\alpha, n)$ a function which converges to a finite value depending on α as n goes to infinity. Define ψ_n to be **the longest first inactivity period in the SB-coalescent process**, given by

$$\psi_n := \max_{1 \leq i \leq M_0 + \mathcal{B}_{N_0}} \xi_i \tag{2.3}$$

where the inactive periods $\xi_i \sim \exp(\sigma)$ are i.i.d.. Additionally define the following processes

$$\underline{\psi}_n := \max_{1 \leq i \leq M_0} \xi_i \quad \text{and} \quad \overline{\psi}_n := \max_{1 \leq i \leq M_0 + \mathcal{B}_n} \xi_i.$$

By definition, it follows that $\underline{\psi}_n \overset{a.s.}{\leq} \psi_n \overset{a.s.}{\leq} \overline{\psi}_n$. We present some properties for $\underline{\psi}_n, \psi_n$ and $\overline{\psi}_n$ below.

Proposition 2.4. *Let* $\widetilde{\psi}_n \in \left\{\underline{\psi}_n, \psi_n, \overline{\psi}_n\right\}$ *then*

a) *Asymptotic distribution.*

$$\sigma\left(\widetilde{\psi}_n - \frac{\ln np}{\sigma}\right) \xrightarrow[n \to \infty]{d} \widetilde{T}, \tag{2.4}$$

where $\widetilde{T} \sim Gumbel(0, 1)$.

b) *Given* $0 < r < \frac{1}{\sigma}$, *it holds that*

$$\lim_{n \to \infty} \mathbb{P}\left[\widetilde{\psi}_n \leq r \ln np\right] = 0. \tag{2.5}$$

c) The sequence $\{\widetilde{\psi}_n\}_{n \in \mathbb{N}}$ almost surely diverges.

Proof. We start proving the proposition for the case $\widetilde{\psi}_n = \overline{\psi}_n$.

a) Given $t > 0$ and $\varepsilon > 0$, using Tchebyshev's inequality we have that

$$\lim_{n \to \infty} \mathbb{P}\left[M_0 + \mathcal{B}_n < (p+\varepsilon)n\right] = 1 \quad \text{and} \quad \lim_{n \to \infty} \mathbb{P}\left[M_0 + \mathcal{B}_n > (p-\varepsilon)n\right] = 1 \quad (2.6)$$

We can write

$$\lim_{n \to \infty} \mathbb{P}\left[\overline{\psi}_n \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}\right] = \lim_{n \to \infty} \mathbb{P}\left[A_n\right],$$

where $A_n = \left\{\overline{\psi}_n \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}; (p-\varepsilon)n \leq M_0 + \mathcal{B}_n \leq (p+\varepsilon)n\right\}$.

Using the definition of $\overline{\psi}_n$ we have that

$$\mathbb{P}\left[\max_{0 \leq i \leq (p+\varepsilon)n} \xi_i \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}\right] \leq \mathbb{P}\left[A_n\right] \leq \mathbb{P}\left[\max_{0 \leq i \leq (p-\varepsilon)n} \xi_i \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}\right],$$

and also

$$\mathbb{P}\left[\max_{0 \leq i \leq (p-\varepsilon)n} \xi_i \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}\right] = \left(1 - \frac{e^{-t}}{np}\right)^{(p-\varepsilon)n} \xrightarrow[n \to \infty]{} \exp\left[-\frac{(p-\varepsilon)e^{-t}}{p}\right].$$

Analogously,

$$\mathbb{P}\left[\max_{0 \leq i \leq (p+\varepsilon)n} \xi_i \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}\right] \xrightarrow[n \to \infty]{} \exp\left[-\frac{(p+\varepsilon)e^{-t}}{p}\right],$$

and letting ε go to zero we conclude that

$$\lim_{n \to \infty} \mathbb{P}\left[\overline{\psi}_n \leq \frac{t}{\sigma} + \frac{\ln np}{\sigma}\right] = \exp(-e^{-t}).$$

The final observation is that $e^{-e^{-t}}$ corresponds to the cumulative distribution function of a standard Gumbel distribution.

b) Using equation (2.3) we have that

$$\begin{aligned}
\mathbb{P}\left[\overline{\psi}_n \leq r \ln np\right] &= \mathbb{P}\left[\overline{\psi}_n \leq r \ln np, \mathcal{B}_n + M_0 > (p-\varepsilon)n\right] \\
&\quad + \mathbb{P}\left[\overline{\psi}_n \leq r \ln np, \mathcal{B}_n + M_0 \leq (p-\varepsilon)n\right] \\
&\leq \mathbb{P}\left[\max_{0 \leq i \leq (p-\varepsilon)n} \xi_i \leq r \ln np\right] + \mathbb{P}\left[\mathcal{B}_n + M_0 \leq (p-\varepsilon)n\right],
\end{aligned}$$

and thanks to equation (2.6) for the case $\varepsilon \in (0, p)$, and the fact that $\sigma r < 1$ by hypothesis, we conclude that

$$\mathbb{P}\left[\overline{\psi}_n \leq r \ln np\right] \leq \left(1 - \frac{1}{(np)^{\sigma r}}\right)^{(p-\varepsilon)n} + \mathbb{P}\left[\mathcal{B}_n + M_0 \leq (p-\varepsilon)n\right] \xrightarrow[n \to \infty]{} 0.$$

c) Given $M \geq 0$ and $r \in (0, \frac{1}{\sigma})$, there exists $K \in \mathbb{N}$ such that $M \leq r \ln np$ for all $n \geq K$. We have then that

$$\lim_{n \to \infty} \mathbb{P}[\overline{\psi}_n \leq M] \leq \lim_{n \to \infty} \mathbb{P}[\overline{\psi}_n \leq r \ln np] = 0,$$

so $\overline{\psi}_n \xrightarrow{\mathbb{P}} \infty$. Since $\overline{\psi}_n \overset{a.s.}{\leq} \overline{\psi}_{n+1}$, we conclude that $\overline{\psi}_n \xrightarrow{a.s.} \infty$.

Note that the core of the proof relies on the equations (2.6). In turn, the equations (2.6) follow from the expected value and variance of $M_0 + \mathcal{B}_n$ given in (2.2). Therefore, an analogous proof for $\underline{\psi}_n$ holds since $\mathbb{E}[M_0] = pn$ and $\mathbb{V}[M_0] = p(1-p)n$. Finally, as $\underline{\psi}_n \overset{a.s.}{\leq} \psi_n \overset{a.s.}{\leq} \overline{\psi}_n$, the result is also true for ψ_n. □

Using this properties, we now study the asymptotic behavior of the time to the most recent common ancestor.

Theorem 2.5. *Given* $(\Pi^n_t)_{t\geq 0}$ *the seed-bank coalescent starting with* n *individuals associated with the Moran lookdown seed-bank model, then*

$$\frac{\widetilde{T}_{MRCA}[n]}{\psi_n} \xrightarrow[n\to\infty]{\mathbb{P}} 1.$$

Moreover, if $\sigma < (1-p)^2$ *then*

$$\widetilde{T} \leq \lim_{n\to\infty} \sigma\left(\widetilde{T}_{MRCA}[n] - \frac{1}{\sigma}\ln np\right) \leq \sigma\varrho + \widetilde{T} \qquad a.s.,$$

where $\widetilde{T} = \lim_{n\to\infty} \sigma\left(\psi_n - \frac{1}{\sigma}\ln np\right) \sim Gumbel(0,1)$ *as in Proposition 2.4 and* ϱ *is the time to the most recent common ancestor of the Kingman coalescent.*

Remark 2.6. The condition $\sigma < (1-p)^2$ corresponds to the case where the rate at which dormant individuals become active is lower than the rate at which a successful reproduction event between two individuals occurs in stationarity.

Corollary 2.7. *The* $\widetilde{T}_{MRCA}[n]$ *and its expected value satisfy*

$$\frac{\widetilde{T}_{MRCA}[n]}{\frac{1}{\sigma}\ln np} \xrightarrow[n\to\infty]{\mathbb{P}} 1 \qquad and \qquad \lim_{n\to\infty} \frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\frac{1}{\sigma}\ln np} = 1, \qquad \forall\alpha,\sigma > 0.$$

Remark 2.8. Corollary 2.7 represents a notable improvement to [1, Theorem 4.6] as it explicitly indicates the long time behavior of the expected value and the first order.

2.2.1 Proof of Theorem 2.5

To begin, note that $\psi_n \overset{a.s.}{\leq} \widetilde{T}_{MRCA}[n]$, since ψ_n corresponds to an inactive period contained in the seed-bank coalescent. So, it is enough to prove that

$$\frac{\widetilde{T}_{MRCA}[n] - \psi_n}{\psi_n} \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

To this end, we will decompose the interval $[0, \widetilde{T}_{MRCA}[n]]$ into three different phases; the Kingman phase, the longest inactive phase and the last coalescent phase. For the Kingman phase we consider the time until ϱ_n, and we know that $\mathbb{E}[\varrho_n] \leq 2$ by [1].

On the other hand, let's analyze the number of lineages after the *longest inactive phase*, the time interval between ϱ_n and ψ_n. Let C_n be the number of levels with lineages that did not coalesce with the first $\ln(np)$ lineages before ψ_n. We will show that C_n goes to zero almost surely when n goes to infinity. In fact, using Remark 2.3, we have that

$$C_n := \sum_{i=\ln np}^{M_0+\mathcal{B}_{N_0}} \mathbb{1}_{\left\{\psi_n < \tau^i_{\ln np}\right\}} \leq \sum_{i=\ln np}^{M_0+\mathcal{B}_n} \mathbb{1}_{\left\{\underline{\psi}_n < \tau^i_{\ln np}\right\}} \leq \sum_{i=\ln np}^{M_0+\mathcal{B}_n} \mathbb{1}_{\left\{\underline{\psi}_n < r^i_{\ln np}\right\}}$$

and recalling definition (2.3), taking conditional expectation with respect to $\{M_0 = m, \mathcal{B}_n = n'\}$ for any $m + n' > \ln np$ we have that

$$
\mathbb{E}\left[\sum_{i=\ln np}^{M_0+\mathcal{B}_n} \mathbb{1}_{\{\underline{\psi}_n < r_{\ln np}^i\}}\middle| M_0 = m, \mathcal{B}_n = n'\right]
$$

$$
= (m + n' - \ln np)\mathbb{E}\left[\mathbb{1}_{\{\underline{\psi}_n < r_{\ln np}^{\ln np+1}\}}\middle| M_0 = m, \mathcal{B}_n = n'\right]
$$

$$
\leq (m + n')\mathbb{P}\left[\max_{1\leq j\leq m}\xi_j < r_{\ln np}^{\ln np+1}\right]
$$

$$
= (m + n')\int_0^\infty \mathbb{P}\left[z < r_{\ln np}^{\ln np+1}\right](1 - e^{-\sigma z})^{m-1}\sigma m e^{-\sigma z}dz
$$

$$
\leq (m + n')\int_0^\infty e^{-(\ln np)(1-p)^2 z}(1 - e^{-\sigma z})^{m-1}\sigma m e^{-\sigma z}dz
$$

$$
= (m + n')\int_0^\infty (np)^{-(1-p)^2 z}(1 - e^{-\sigma z})^{m-1}\sigma m e^{-\sigma z}dz
$$

$$
= (m + n')\mathbb{E}\left[(np)^{-(1-p)^2 \max_{1\leq j\leq m}\xi_j}\right]
$$

$$
= (m + n')\mathbb{E}\left[(np)^{-(1-p)^2\underline{\psi}_n}\middle| M_0 = m\right]
$$

$$
= \mathbb{E}\left[(M_0 + \mathcal{B}_n)(np)^{-(1-p)^2\underline{\psi}_n}\middle| M_0 = m, \mathcal{B}_n = n'\right].
$$

The penultimate equality comes from the fact that $\underline{\psi}_n$ conditioned to $M_0 = m$ is distributed as the maximum of m exponential random variables of parameter σ whose density is given by $(1 - e^{-\sigma z})^{m-1}\sigma m e^{-\sigma z}$. Now, using L_2-Cauchy inequality and the fact that $M_0 + \mathcal{B}_n \geq (p - \varepsilon)n > \ln np$ almost surely for n sufficiently large by equation (2.6), we have that

$$
\mathbb{E}[C_n] \leq \mathbb{E}\left[\sum_{i=\ln np}^{M_0+\mathcal{B}_n} \mathbb{1}_{\{\underline{\psi}_n < r_{\ln np}^i\}}\right] \leq \mathbb{E}\left[\mathbb{E}\left[(M_0 + \mathcal{B}_n)(np)^{-(1-p)^2\underline{\psi}_n}\middle| M_0, \mathcal{B}_n\right]\right]
$$

$$
\leq \mathbb{E}\left[\mathbb{E}\left[(M_0 + \mathcal{B}_n)(np)^{-(1-p)^2\underline{\psi}_n}\middle| M_0, \mathcal{B}_n\right]\right] = \mathbb{E}\left[(M_0 + \mathcal{B}_n)(np)^{-(1-p)^2\underline{\psi}_n}\right]
$$

$$
\leq \left(\mathbb{V}(M_0 + \mathcal{B}_n) + \mathbb{E}^2[M_0 + \mathcal{B}_n]\right)^{\frac{1}{2}}\mathbb{E}\left[(np)^{-2(1-p)^2\underline{\psi}_n}\right]^{\frac{1}{2}}
$$

$$
\leq \left(\mathbb{E}[M_0 + \mathcal{B}_n] + \mathbb{E}^2[M_0 + \mathcal{B}_n]\right)^{\frac{1}{2}}\mathbb{E}\left[(np)^{-2(1-p)^2\underline{\psi}_n}\right]^{\frac{1}{2}}
$$

$$
\leq \left(np + 2\alpha\ln n + R(\alpha, n) + (np + 2\alpha\ln n + R(\alpha, n))^2\right)^{\frac{1}{2}}\mathbb{E}\left[(np)^{-2(1-p)^2\underline{\psi}_n}\right]^{\frac{1}{2}},
$$

where in the last line we use equation (2.2). Since $\{\underline{\psi}_n\}_{n\in\mathbb{N}}$ diverges almost surely, $\lim_{n\to\infty}\mathbb{E}[(np)^{k-2(1-p)^2\underline{\psi}_n}] = 0$ for any $k \geq 0$, so we can deduce that

$$
\lim_{n\to 0}(pn)^k\mathbb{E}[C_n] = 0. \tag{2.7}
$$

Moreover, as C_n is non-negative, we conclude that $\lim_{n\to\infty} C_n \overset{a.s.}{=} 0$. We will employ the same argument to check that $\widetilde{T}_{MRCA}[\ln np + C_n] - \widetilde{T}_{MRCA}[\ln np] \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Let $\varepsilon > 0$, using

Markov inequality we can deduce that

$$\mathbb{P}\left[\widetilde{T}_{MRCA}[\ln np + C_n] - \widetilde{T}_{MRCA}[\ln np] \geq \varepsilon\right] \leq \frac{1}{\varepsilon}\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]\mathbb{P}[C_n \geq 1]$$

$$\leq \frac{1}{\varepsilon}\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]\mathbb{E}[C_n]$$

$$\leq \frac{1}{\epsilon}\frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\ln np}\ln np\,\mathbb{E}[C_n]$$

by [1, Remark 4.13] we know that

$$\limsup_{n\to\infty}\frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\ln np} < \infty, \tag{2.8}$$

whereas $\ln(np)\mathbb{E}[C_n]$ goes to 0 by (2.7).

Therefore, $\widetilde{T}_{MRCA}[n]$ is bounded by

$$\psi_n \overset{a.s.}{\leq} \widetilde{T}_{MRCA}[n] \overset{a.s.}{\leq} \varrho_n + \psi_n + \widetilde{T}_{MRCA}[\ln np + C_n] \tag{2.9}$$

and from here we have that

$$\frac{\widetilde{T}_{MRCA}[n] - \psi_n}{\psi_n} \overset{a.s.}{\leq} \frac{\varrho_n + \widetilde{T}_{MRCA}[\ln np + C_n] - \widetilde{T}_{MRCA}[\ln np]}{\psi_n} + \frac{\widetilde{T}_{MRCA}[\ln np]}{\psi_n}. \tag{2.10}$$

For any $\varepsilon > 0$ and $r \in (0, \frac{1}{\sigma})$ we have

$$\mathbb{P}\left[\frac{\widetilde{T}_{MRCA}[\ln(np)]}{\psi_n} > \varepsilon\right] = \mathbb{P}\left[\widetilde{T}_{MRCA}[\ln(np)] > \varepsilon\psi_n,\,\psi_n \leq r\ln(np)\right]$$

$$+ \mathbb{P}\left[\widetilde{T}_{MRCA}[\ln(np)] > \varepsilon\psi_n,\,\psi_n > r\ln(np))\right]$$

$$\leq \mathbb{P}\left[\psi_n \leq r\ln(np)\right] + \mathbb{P}\left[\widetilde{T}_{MRCA}[\ln(np)] > \varepsilon r\ln(np)\right]$$

$$\leq \mathbb{P}\left[\psi_n \leq r\ln(np)\right] + \frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[\ln(np)]\right]}{\varepsilon r\ln(np)}.$$

The first term goes to zero according with Proposition (2.4). We will show that the second term also goes to zero via the generalized dominated convergence theorem, thanks to (2.8), it follows that

$$\lim_{n\to\infty}\frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[\ln(np)]\right]}{\varepsilon r\ln(np)} = 0. \tag{2.11}$$

Finally, from (2.10) we conclude that $\frac{\widetilde{T}_{MRCA}[n]-\psi_n}{\psi_n}\xrightarrow[n\to\infty]{\mathbb{P}}0$. For the remaining part of the proof we study those lineages that have not coalesced with the lineage of the first level individual once the longest inactivity period concludes. Note that

$$\widetilde{C}_n = \sum_{i=2}^{n}\mathbb{1}_{\left\{\psi_n \leq \tau_1^i\right\}} \leq \sum_{i=2}^{n}\mathbb{1}_{\left\{\psi_n \leq r_1^i\right\}}$$

which leads to a similar inequality of (2.9)

$$\psi_n \overset{a.s.}{\leq} \widetilde{T}_{MRCA}[n] \overset{a.s.}{\leq} \rho_n + \psi_n + \widetilde{T}_{MRCA}[1 + \widetilde{C}_n]. \tag{2.12}$$

Using conditional expected value

$$\mathbb{E}\left[\widetilde{C}_n\Big|\psi_n\right] = \sum_{i=2}^{n}\mathbb{E}\left[\mathbb{1}_{\{\psi_n\leq\tau_1^i\}}\Big|\psi_n\right] \leq \sum_{i=2}^{n}\mathbb{E}\left[\mathbb{1}_{\{\psi_n\leq r_1^i\}}\Big|\psi_n\right] \leq ne^{-(1-p)^2\psi_n}$$

then

$$\mathbb{E}\left[\widetilde{C}_n\right] = \mathbb{E}\left[\mathbb{E}\left[\widetilde{C}_n\Big|\psi_n\right]\right] \leq n^{1-\frac{(1-p)^2}{\sigma}}p^{-\frac{(1-p)^2}{\sigma}}\mathbb{E}\left[e^{-\frac{(1-p)^2}{\sigma}\cdot\sigma\left(\psi_n-\frac{1}{\sigma}\ln np\right)}\right].$$

By Lemma 2.4.a the expected value in the right hand side converges to the moment generating function of the Gumbel distribution. Additionally, if $\sigma < (1-p)^2$ then $\mathbb{E}\left[\widetilde{C}_n\right]$ converges to 0 as n approaches infinity. Consequently, $\widetilde{C}_n$ converges almost surely to 0 since it is positive, enabling us to conclude that

$$\lim_{n\to\infty}\widetilde{T}_{MRCA}[1+\widetilde{C}_n] = \widetilde{T}_{MRCA}\left[1+\lim_{n\to\infty}\widetilde{C}_n\right] \overset{a.s.}{=} 0.$$

Finally, rescaling (2.12) we conclude the proof where ϱ correspond to the time to the most recent common ancestor in the Kingman coalescent. $\square$

Proof of Corollary 2.7. By Slutsky's theorem to deduce that

$$\frac{\psi_n}{\frac{1}{\sigma}\ln np} = \frac{\psi_n-\frac{1}{\sigma}\ln np}{\frac{1}{\sigma}\ln np}+1 \xrightarrow[n\to\infty]{\mathbb{P}} 1$$

which implies that $\frac{\widetilde{T}_{MRCA}[n]}{\frac{1}{\sigma}\ln np} \xrightarrow[n\to\infty]{\mathbb{P}} 1$. Using the left side of the identity (2.9) and Proposition 2.4.b for $r < \frac{1}{\sigma}$ it follows that

$$\sigma r = \lim_{n\to\infty}\sigma r\mathbb{P}\left[r\ln np\leq\psi_n\right] \leq \lim_{n\to\infty}\frac{\mathbb{E}[\psi_n\mathbb{1}_{\{r\ln np\leq\psi_n\}}]}{\frac{1}{\sigma}\ln np} \leq \lim_{n\to\infty}\frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\frac{1}{\sigma}\ln np}.$$

By letting r approach to $\frac{1}{\sigma}$ we deduce that $1 \leq \lim_{n\to\infty}\frac{\mathbb{E}[\widetilde{T}_{MRCA}[n]]}{\frac{1}{\sigma}\ln np}$. We will use the right hand side of the identity (2.9) to proof the upper bound inequality. Since $\mathbb{E}[\varrho_n]\leq 2$, we have that $\lim_{n\to\infty}\frac{\mathbb{E}[\varrho_n]}{\frac{1}{\sigma}\ln np} = 0$. Note that

$$\frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[\ln np + C_n]\right]}{\ln np} \leq \frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\ln np}\mathbb{P}\left[C_n\geq 1\right] + \frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[\ln np]\right]}{\ln np}$$

which goes to 0 as n tends to infinity by (2.7), (2.8) and (2.11). Finally, since ψ_n is bounded by the maximum of n independent exponential random variables of rate σ we have that

$$\lim_{n\to\infty}\frac{\mathbb{E}\left[\widetilde{T}_{MRCA}[n]\right]}{\frac{1}{\sigma}\ln np} \leq \lim_{n\to\infty}\frac{\mathbb{E}\left[\psi_n\right]}{\frac{1}{\sigma}\ln np} \leq \lim_{n\to\infty}\frac{\mathbb{E}\left[\max_{1\leq i\leq n}\xi_i\right]}{\frac{1}{\sigma}\ln np} = \lim_{n\to\infty}\frac{\sum_{i=1}^{n}\frac{1}{i}}{\ln np} = 1$$

which concludes the proof. $\square$

3 Fixation time of a type in the SB-lookdown

Suppose we start with an initial exchangeable distribution $\left(Z_1^N(0),\ldots,Z_N^N(0)\right)$ which sets at most one individual per type. Let $\widetilde{\tau}^N$ be the fixation time of type in the first level for the N-lookdown model, given by

$$\widetilde{\tau}^N := \inf\left\{t\geq 0 : Z_i^{N,E}(t) = Z_1^{N,E}(0), i\in\{1,2,\cdots,N\}\right\}. \tag{3.1}$$

If we study the the time to the most recent common ancestor $\widetilde{T}_{MRCA}[N]$ from a fixed initial time $\zeta \geq 0$, we will see that there are only the following two scenarios.

1. **Scenario 1.** If $\widetilde{T}_{MRCA}[N] \leq \zeta$, this implies that $(\zeta - \widetilde{T}_{MRCA}[N], \zeta) \subseteq (0, \zeta)$, and the fixation time $\widetilde{\tau}^N$ also occurs before ζ, since each individual has already coalesced with the individual in the first level.

2. **Scenario 2.** If $\widetilde{T}_{MRCA}[N] > \zeta$, this implies that $(0, \zeta) \subset (\zeta - \widetilde{T}_{MRCA}[N], \zeta)$ and the fixation time $\widetilde{\tau}^N$ is bigger than ζ, because at time zero the individual in the first level has not yet coalesced with all the $N - 1$ remaining individuals.

From before, we deduce that $\mathbb{P}\left[\widetilde{\tau}^N \leq \zeta\right] \leq \mathbb{P}\left[\widetilde{T}_{MRCA}[N] \leq \zeta\right]$ and $\mathbb{P}\left[\widetilde{T}_{MRCA}[N] \leq \zeta\right] \leq \mathbb{P}\left[\widetilde{\tau}^N \leq \zeta\right]$, so $\widetilde{\tau}^N \stackrel{d}{=} \widetilde{T}_{MRCA}[N]$.

Using this equality in law, Theorem 2.5 and Corollary 2.7, we obtain the first order of the fixation time as well as its distribution approximation under the case $\sigma < (1-p)^2$.

Corollary 3.1. *Given $\widetilde{\tau}^N$ the fixation time of the type in the first level for the N-lookdown model, it holds that*

$$\frac{\widetilde{\tau}^N}{\frac{1}{\sigma}\ln Np} \xrightarrow[n\to\infty]{\mathbb{P}} 1 \qquad and \qquad \lim_{N\to\infty} \frac{\mathbb{E}\left[\widetilde{\tau}^N\right]}{\frac{1}{\sigma}\ln Np} = 1, \qquad \forall \alpha, \sigma > 0.$$

Moreover if $\sigma < (1-p)^2$ then for all $t \in \mathbb{R}$

$$\mathbb{P}\left[\sigma\varrho + \widetilde{T} \leq t\right] \leq \mathbb{P}\left[\lim_{N\to\infty} \sigma\left(\widetilde{\tau}^N - \frac{1}{\sigma}\ln Np\right) \leq t\right] \leq \mathbb{P}\left[\widetilde{T} \leq t\right]$$

where $\widetilde{T} \sim Gumbel(0,1)$ and ϱ is distributed as the time to the most recent common ancestor of the Kingman coalescent.

References

[1] Jochen Blath, Adrián González Casanova, Noemi Kurt, and Maite Wilke-Berenguer, *A new coalescent for seed-bank models*, The Annals of Applied Probability **26** (2016), no. 2, 857–891. MR3476627

[2] Peter Donnelly and Thomas G. Kurtz, *A countable representation of the Fleming-Viot measure-valued diffusion*, The Annals of Probability **24** (1996), no. 2, 698 – 742. MR1404525

[3] Peter Donnelly and Thomas G. Kurtz, *Particle Representations for Measure-Valued Population Models*, The Annals of Probability **27** (1999), no. 1, 166 – 205. MR1681126

[4] Alison M. Etheridge and Thomas G. Kurtz, *Genealogical constructions of population models*, The Annals of Probability **47** (2019), no. 4, 1827 – 1910. MR3980910

[5] Adrián González Casanova, Lizbeth Peñaloza, and Arno Siri-Jégousse, *The shape of a seed bank tree*, Journal of Applied Probability **59** (2022), no. 3, 631–651. MR4480072

[6] Andreas Greven, Frank den Hollander, and Margriet Oomen, *Spatial populations with seed-bank: well-posedness, duality and equilibrium*, Electronic Journal of Probability **27** (2022), no. none, 1 – 88. MR4377129

[7] Ingemar Kaj, Stephen Krone, and Martin Lascoux, *Coalescent theory for seed bank models*, Journal of Applied Probability **38** (2001), 285–300. MR1834743

[8] Thomas G. Kurtz, *Martingale Problems for Conditional Distributions of Markov Processes*, Electronic Journal of Probability **3** (1998), no. none, 1 – 29. MR1637085

[9] Thomas G. Kurtz and Eliane R. Rodrigues, *Poisson representations of branching Markov and measure-valued branching processes*, The Annals of Probability **39** (2011), no. 3, 939 – 984. MR2789580