Moran model with strong selection and Λ-Wright-Fisher SDE

François Gaston Ged

Abstract

We study a population model of fixed size undergoing strong selection where individuals accumulate beneficial mutations, namely the Moran model with selection. In his specific setting of [15, 16], Schweinsberg showed that, due to the strong selection, the genealogy of the population is described by the so-called Bolthausen-Sznitman’s coalescent. In this paper we sophisticate the model by splitting the population into two adversarial subgroups, that can be interpreted as two different alleles, one of which has a selective advantage over the other. We show that the proportion of disadvantaged individuals converges to the solution of a stochastic differential equation (SDE) as the population’s size goes to infinity, named the Λ-Wright-Fisher SDE with selection. This stochastic differential equation already appeared in the Λ-lookdown model with selection studied by Bah and Pardoux [1], in the case where the population’s genealogy is described by Bolthausen-Sznitman’s coalescent.

Subject classification: 60J80, 92D15, 92D25, 60H10.

Keywords: Moran model with selection, Bolthausen-Sznitman’s coalescent, Λ-Wright-Fisher SDE.

1 Introduction

The Moran model is a classical model in population genetics. It describes the evolution in continuous time of a haploid population with constant size, where generations are overlapping. Every individual dies at rate 1 and is instantaneously replaced by a copy of an individual chosen uniformly at random in the remaining population, including the individual who just died. It is well-known that the genealogy of the Moran model is described by the so-called Kingman’s coalescent, which is the only exchangeable coalescent process where merging events are only binary and non-simultaneous. Some kind of universality of Kingman’s coalescent for the genealogy of discrete time population models with fixed size was established in [12]; this result is known as Möhle’s Lemma. In [13], Möhle also obtained convergence results towards different coalescents, and even allowed the size of the population to vary.

In the Moran model of size $N$, when the population is split into two subgroups, say the individuals carrying allele $B$ and the ones carrying allele $b$, the proportion $(X_t)_{t \geq 0}$ of allele $B$ in the whole population converges as $N \to \infty$, when speeding up the time by a factor $N$, towards the Wright-Fisher diffusion, that is the solution to the SDE

$$dX_t = \frac{1}{2} \sqrt{X_t(1-X_t)} dW_t,$$

where $dW_t$ is a Wiener process. This SDE is known as the Wright-Fisher diffusion with selection.

1 Exchangeable refers to the property that permuting two individuals leaves the law of the process unchanged.
where $W$ is a standard Brownian motion. Note the symmetry between the two alleles reflecting the fact that none of them has a selective advantage over the other. One of the implications is the well-known duality relation between the number of blocks in Kingman’s coalescent and the Wright-Fisher diffusion, as stated in Theorem 2.7 in [2]. Namely, denoting $K_t$ the number of blocks in Kingman’s coalescent at time $t$, it holds for all $x \in (0,1)$ and $k \in \mathbb{N}$ that
\[
E(X^x_t | X_0 = x) = E(x^{K_t} | K_0 = k).
\] (1)

The duality actually holds for more general coalescents, namely $\Lambda$-coalescents, and some Fleming-Viot processes, as shown in [3] equation (18). The coalescence rates of a $\Lambda$-coalescent are characterized by a finite measure $\Lambda$ on $[0,1]$, such that if the coalescent contains $k$ block at some given time, any sub-family of size $\ell$ among the $k$ blocks merge at rate given by
\[
\lambda_{k,\ell} := \int_{[0,1]} p^{\ell-2} (1 - p)^{k-\ell} \Lambda(dp).
\]

In particular, the blocks are exchangeable, in the sense that all the possible combinations of $\ell$ blocks have the same rate of merging. Kingman’s coalescent corresponds to the case $\Lambda = \delta_0$, the Dirac mass at 0. Another instance of $\Lambda$-coalescent is the Bolthausen-Sznitman coalescent introduced in [6], corresponding to $\Lambda(dp) = dp$. Its importance is due to its connections to models such as spin glasses, continuous branching processes, travelling waves, some population models, see e.g. [2] and references therein. The populations where we expect to observe Bolthausen-Sznitman coalescent are for instance populations undergoing strong selection [16], exploring uninhabited territories [7], or quickly adapting to the environment [14, 9]. In those cases, an individual sometimes reproduces more (or faster) and generates a family of size of the same order as the population size.

**Moran model with selection and $\Lambda$-lookdown model.** When a death occurs in the Moran model, instead of choosing an individual uniformly at random to reproduce, we can include a selection component in the dynamics and choose the parent proportionally to its fitness. An instance of a Moran model with selection has been studied by Schweinsberg in [15] and [16], where the individuals accumulate beneficial mutations increasing their reproduction rates; this is the model we are interested in in this work and we will describe it in more details later on. The main result of [16] establishes that the genealogy of the Moran model with selection of Schweinsberg converges towards Bolthausen-Sznitman coalescent, as the size of the population goes to infinity. Let us connect it with another population model.

In [1], the authors study an infinite size population model called the $\Lambda$-lookdown model with selection, whose genealogy is that of the corresponding $\Lambda$-coalescent. We are of course interested in the Bolthausen-Sznitman case $\Lambda(dp) = dp$ on $[0,1]$. Each individual carries either allele $B$ or allele $b$, the selection advantaging the carriers of $B$. Theorem 3.5 in [1] shows that the proportion of carriers of $b$ is the solution of the following SDE:
\[
\mathcal{Y}_t = \mathcal{Y}_0 - \alpha \int_0^t \mathcal{Y}_s (1 - \mathcal{Y}_s) ds + \int_{[0,t] \times [0,1]^2} p(1_{u \leq \mathcal{Y}_s} - \mathcal{Y}_s) M(ds, du, dp),
\] (2)

where $M$ is a Poisson point process with intensity $dt \otimes du \otimes dp$, and $\alpha \geq 0$ represents the selective advantage of $B$ over $b$. In [1], Equation (2) is called the $\Lambda$-Wright-Fisher SDE with selection. In view of the previously mentionned duality between the coalescent and
the proportion process of a population, one may wonder whether it is possible to split the individuals in the Moran model with selection into two adversarial subgroups (B versus b), in order to observe the convergence of the proportion of the disadvantaged group b towards the solution of (2). The goal of this work is to answer this question.

2 Model and main result

2.1 Previous results

We describe more formally the Moran model with selection and the results of Schweinsberg in [15] and [16].

We consider a population of fixed size $N \in \mathbb{N}$. Each individual dies at rate 1, meaning that its lifetime is an exponential random variable with parameter 1. At time 0, the $N$ individuals carry no mutation. Each of them acquires a mutation that adds up to its current number of mutation, at rate $\mu = \mu_N$ that can depend on $N$. We call the number of mutations carried by an individual its type. When a death occurs, say at time $t$, the individual is instantaneously replaced by a copy of an individual chosen in the population at time $t$, including the one who just died, independently from the past. The parent is chosen at random proportionally to its fitness at time $t$, as explained below, and the newborn individual then inherits the type of its parent.

For all $j \geq 0$ and $t \in \mathbb{R}_+$, we denote by $W_j(t)$ the number of individuals of type $j$ at time $t$ in the population. The average number of mutations at time $t$ is thus given by

$$M(t) := \frac{1}{N} \sum_{j \geq 0} jW_j(t).$$

Let $s = s_N > 0$ be the coefficient of selection and let the fitness of the type $j$ at time $t$ be $\max(1 + s(j - M(t)), 0)$. If a death occurs at time $t$, the probability that a particular individual of type $j$ reproduces is

$$F_j(t) := \frac{\max(1 + s(j - M(t)), 0)}{\sum_{i \geq 0} W_i(t) \max(1 + s(i - M(t)), 0)},$$

which becomes

$$\frac{1 + s(j - M(t))}{N}$$

when all the fitnesses are positive. The neutral Moran model corresponds to the case where $s = 0$, that is all the individuals have the same probability to reproduce. Define

$$k_N := \frac{\log N}{\log(s/\mu)} \quad \text{and} \quad a_N := \frac{\log(s/\mu)}{s}, \quad (3)$$

which are proven in [15] to be the scaling constants such that in $a_N$ units of time, the number of new types appearing is of order $k_N$. The assumptions on the parameters of the model are the following:

$$(A_1) : \lim_{N \to \infty} \frac{k_N}{\log(1/s)} = \infty.$$
\[(A_2) : \lim_{N \to \infty} \frac{k_N \log k_N}{\log(s/\mu)} = 0.\]

\[(A_3) : \lim_{N \to \infty} sk_N = 0.\]

In particular, it implies that \(s \to 0, k_N, a_N \to \infty\) as \(N \to \infty\), and for any \(a, b > 0\),

\[
\frac{1}{N^a} \ll \mu \ll s^b. \tag{4}
\]

We refer to [15] and [16] for more detailed discussions on these assumptions.

The main results of Schweinsberg in [15] concern the dynamics of the types distribution in the population as \(N \to \infty\). Theorem 1.4 in [15] shows that after \(a_N\) units of time, the distribution of the types starts looking like that of a Gaussian variable with small variance. Theorem 1.2 in [15] states that \(M(a_N t)/k_N\) converges in probability and uniformly on compact sets of \((0, 1) \cup (1, \infty)\) towards a function \(m\) that we do not need to describe here. A similar convergence holds for the difference between the fittest individuals (the highest type alive) and the mean type, as shown in Theorem 1.1 in [15]. This fully describes the dynamics of the types distribution as \(N \to \infty\) forward in time. It enabled Schweinsberg in [16] to show, when looking backward in time, that the genealogy of the process converges in finite distributions towards the Bolthausen-Sznitman coalescent.

**Following the leading type.** An important observation in [16] is that, when sampling \(n\) individuals in the population at time \(a_N T\) and looking backwards in time, after \(a_N\) units of time, all the individuals essentially share the same type with high probability and it goes on for their ancestors, the common type being the leading type in the population. In other words, after \(a_N\) units of time forward, only the individuals that were leading have begotten a non-negligible offsprings. Hence, Schweinsberg discretises the time at stopping times defined as follows: for all \(j \geq 1\), let

\[
\tau_j := \inf \{ t \geq 0 : W_{j-1}(t) > s/\mu \}. \tag{5}
\]

In words, \(\tau_j\) is approximately the time after which type \(j\) mutations start occuring, making \(j - 1\) the fittest type in the population at time \(\tau_j\); see [15] Equation (3.16) and the associated discussion. Note that \(s/\mu \to \infty\) as \(N \to \infty\) by (4), but \(s/\mu N \to 0\). Roughly speaking, it means that although the leading type represents a proportion of the whole population close to 0, once it reaches a size \(\lceil s/\mu \rceil := \inf \{ n \geq 1 : n \geq s/\mu \}\), it starts evolving in a very predictable way, which is the reason why this discretisation is powerful. In particular, it is when a type \(j\) mutation occurs relatively shortly after \(\tau_j\) that a large family is likely to descent from it, due to the fact that the fitness is relative to the mean (meaning that individuals mutating faster than usual are getting strongly advantaged for reproducing). To follow the leading type, we introduce the index

\[
j(t) := \sup \{ j \geq 1 : \tau_j \leq a_N t \} \tag{6}\]

**2.2 An additional competition**

Recall that we want to divide our population into two adversarial subgroups, say \(X\) and \(Y\), giving a selective advantage to \(X\) such that the proportion of \(Y\)-individuals converges towards
the solution of (2) as \( N \to \infty \). It is important to note that this new selection is a competition between groups \( X \) and \( Y \), leaving unchanged the selection between the different types.

For technical reasons due to the fact that the population takes about \( a_N \) units of time to reach the Bolthausen-Sznitman dynamics, we study the proportion of \( Y \)-individuals starting only from time \( \tau_{j(2)} \), when the types distribution already looks like a Gaussian distribution.

Let \((y_N)_{N \in \mathbb{N}}\) be a sequence in \((0, 1)\) such that \( y_N \to y \in (0, 1) \) as \( N \to \infty \). At time \( \tau_{j(2)} \), we mark \( y_N \lceil s/\mu \rceil \) type \( j(2) - 1 \) individuals to be in the group \( Y \). All the others individuals form the group \( X \). We add a new selection operating between groups \( X \) and \( Y \) as follows. Set the selection coefficient \( \alpha \geq 0 \) and \( N \) large enough such that \( \alpha < \frac{k}{2a_N} \). Define \( \Delta_j := \frac{\tau_{j+1} - \tau_j}{a_N} \). Let \( Y_j(\tau_{j+1}) \) be the number of \( Y \)-individuals of type \( j \) at time \( \tau_{j+1} \) for \( j \geq j(2) \). At each step at time \( \tau_{j+1} \), we kill uniformly at random \( \lceil \alpha \Delta_j \frac{Y_j(\tau_{j+1})}{s/\mu} \rceil \) individuals among the \( Y_j(\tau_{j+1}) \) \( Y \)-individuals. Each death is immediately compensated by choosing an individual among the \( X_j(\tau_{j+1}) \) \( X \)-individuals uniformly at random with replacement to give birth. It is somehow natural to choose this form of competition, as it can be interpreted as follows: if at rate \( \alpha \) a meeting between two uniformly sampled individuals occurs, such that if an \( X \)-individual meets a \( Y \)-individual, it kills it and replaces it by a copy of itself (any other configuration resulting in nothing), then the average number of killing after a time lapse \( \Delta_j \) is at distance less than 1 from (7).

Let \((Y^N_t)_{t \geq 2}\) be the càdlàg version of the process which, informally, follows the proportion of \( Y \)-individuals among the fittest ones, that is

\[
Y^N_t := \frac{Y_j(t(\tau_{j(t)}) - 1}{s/\mu}. 
\]

We now state our main result.

**Theorem 1.** Given that \( Y^N_2 = y_N \in (0, 1) \), for all \( T > 2 \), the process \((Y^N_t)_{t \in [2, T]}\) converges weakly in the Skorokhod space towards the unique solution of (2).

The strong uniqueness of solutions of (2) is proven in [8] Theorem 4.1.

**Organisation of the paper.** In Section 3, we provide the technical tools for the proof of Theorem 1. It is divided into three subsections.

The first one recalls the notation, as well as useful results from [15] and [16]. Proposition 1 describes the evolution of the types and related quantities, Lemma 1 controls the time lapse between \( \tau_j \) and \( \tau_{j+1} \).

The second subsection adapts Schweinsberg techniques based on martingales to investigate the fluctuations of the number of **non-early** \( Y \)-individual from stage \( j \) to \( j + 1 \). Lemmas 2 and 3 describe the evolution of type \( j \) \( Y \)-individuals after time \( \tau_{j+1} \), when their growth is expected to be predictable (as for the growth of all individuals of type \( j \)). Then, Lemma 4 bounds the variance of the martingale associated with the non-early individuals.

In the third subsection, Lemma 5 gives a description of the increments of the proportion of \( Y \)-individuals given the proportion of early individuals. Lemma 6 provides bounds on the two
first moments of these increments. Finally, Lemma 7 approximates the law of the proportion of early individuals with the Bolthausen-Sznitman rates.

Section 4 is devoted to the proof of Theorem 1. The strategy is the following. We first establish the tightness of $Y^N$ in Lemma 8. Next, we show in Lemma 9 that the expectation of the increment of the proportion of $Y$-individuals from $j$ to $j+1$ is very close to the generator of the solution of (2). We then introduce a martingale problem in Lemma 10 and show that any weak limit of $Y^N$ solves it. We conclude the argument with Lemma 11 who states that this weak limit is therefore a solution of (2).

3 Toolbox

3.1 Schweinsberg’s setting and notation

In this subsection, we introduce the notation used in [15, 16] and we recall some of the results we will need. Thus, what follows do not directly concern the dynamics of the two groups $X$ and $Y$, but rather that of the types distribution. Set $T > 2$ a positive real number, arbitrary large. Fix $\epsilon, \delta \in (0, 1)$ such that

$$\delta < \min \left\{ \frac{1}{100}, \frac{1}{19T}, \epsilon^3 \right\},$$

as required in [16] Equation (5.1). We will study the process up to time $a_N T$, and control its behaviour with a probability greater than $1 - \epsilon$, with accuracy $\delta$. We shall denote by $C_i, i \in \mathbb{N}$, constants that can depend on $\delta, \epsilon, T$ whereas $C$ will always refer to a constant independent of those parameters, that may vary from line to line.

We introduce some tools to study the evolution of a type. Denote $B_j(t), D_j(t)$ respectively the birth-rate and death-rate of the type $j$ at time $t \leq a_N T$, that is:

$$B_j(t) := (N - W_j(t)) F_j(t),$$

$$D_j(t) := \mu + 1 - W_j(t) F_j(t),$$

and define

$$G_j(t) := B_j(t) - D_j(t).$$

The value of $G_j(t)$ is the growth-rate of the type $j$ at time $t$. Thus, as in [15] and [16], we can define for all $j \geq 0$

$$(Z_j(t))_{t \in [0, a_N T]} := \left( e^{-\int_0^t G_j(v)dv} W_j(t) - \int_0^t \mu W_{j-1}(u) e^{-\int_0^u G_j(v)dv} du - W_j(0) \right)_{t \in [0, a_N T]}.$$

This is a square integrable martingale, the variance of which being given for $t \in [0, a_N T]$ by

$$\text{Var}(Z_j(t)) = \mathbb{E} \left( \int_0^t e^{-2 \int_0^u G_j(v)dv} \left( \mu W_{j-1}(u) + B_j(u) W_j(u) + D_j(u) W_j(u) \right) du \right),$$

see [15] Proposition 5.1. The role of $Z_j(t)$ is to control the fluctuations of $W_j(t)$. Note that for every $t \in [\tau_j, \tau_{j+1}]$ one has

$$2 \leq B_j(t) + D_j(t) \leq 3.$$
where the upper bound actually holds for all \( t \). We will often work with variants of the martingales \( Z_j \). Let \( (\mathcal{F}_t^N)_{t \geq 0} \) denote the natural filtration of \((W_j(t), j \geq 0)_{t \geq 0}\). Using classical arguments on martingales and the fact that \((W_j(t), j \geq 0)_{t \geq 0}\) is a strong Markov process, it is shown in [15] Corollary 5.3 that if \( \sigma < \kappa \) are two stopping times, then \( Z_j^{\sigma, \kappa} \) defined for all \( t \in [\sigma, \kappa] \) by:

\[
Z_j^{\sigma, \kappa}(t) := e^{-\int_{\sigma}^{t \wedge \kappa} G_j(v) dv} W_j(t \wedge \kappa) - \int_{\sigma}^{t \wedge \kappa} \mu W_{j-1}(u) e^{-\int_{\sigma}^{u} G_j(v) dv} du - W_j(\sigma)
\]

is a squared integrable martingale with conditional variance

\[
\text{Var} \left( Z_j^{\sigma, \kappa}(\sigma + t) \bigg| \mathcal{F}_\sigma^N \right) = \mathbb{E} \left( \int_{\sigma}^{(\sigma + t) \wedge \kappa} e^{-2 \int_{\sigma}^{u} G_j(v) dv} \left( \mu W_{j-1}(u) + B_j(u) W_j(u) + D_j(u) W_j(u) \right) du \bigg| \mathcal{F}_\sigma^N \right).
\]

If \( S \) is a subpopulation of individuals of type \( j \) at time \( \sigma \), for \( t \geq \sigma \) we denote by \( W_j^S(t) \) the number of individuals of type \( j \) at time \( t \) such that their ancestor at time \( \sigma \) is in \( S \). Define

\[
B_j^S(u) := (N - W_j^S(u)) F_j(u),
\]

\[
D_j^S(u) := \mu + 1 - W_j^S(u) F_j(u),
\]

and one can define, in the same manner as above (see [16] Corollary 4.9), the martingales \( Z_j^S \), the variance of which being given by

\[
\text{Var} \left( Z_j^S(\sigma + t) \bigg| \mathcal{F}_\sigma^N \right) = \mathbb{E} \left( \int_{\sigma}^{\sigma + t} e^{-2 \int_{\sigma}^{u} G_j(v) dv} \left( B_j^S(u) W_j^S(u) + D_j^S(u) W_j^S(u) \right) du \bigg| \mathcal{F}_\sigma^N \right).
\]

In [15] [16], Schweinsberg often distinguishes whether \( j \) is greater or smaller than \( k^* := [k_N^+ - 1] \), where

\[
k_N^+ := k_N + \frac{2k_N \log k_N}{\log(s/\mu)}.
\]

This constant is, roughly speaking, the first type after which the types distribution looks like a Gaussian distribution. For \( j \geq k^* + 1 \), let

\[
q_j^* := \begin{cases} j - k_N & \text{if } a_N - 2a_N/k_N \leq \tau_j \leq a_N + 2a_N/k_N, \\ j - M(\tau_j) & \text{otherwise}. \end{cases}
\]

Recall that \( \tau_j \) defined in [5] is approximately the time where one expects to see the first type \( j \) mutations, hence \( q_j \) is an approximation of the difference between \( j \) and the average number of mutations when individuals of type \( j \) start appearing. Set

\[
b := \log \frac{24000T}{\delta^2 e}.
\]
For \( j \geq k^* + 1 \), define

\[
\xi_j := \max \left\{ \tau_j, \tau_j + \frac{1}{s q_j} \log \left( \frac{1}{s q_j} \right) + \frac{b}{s q_j} \right\}.
\]  

(17)

We shall work on a specific event, realized with high probability, such that it holds that \( \tau_j < \xi_j < \tau_{j+1} \) for all \( j \geq k^* + 1 \) such that \( \tau_{j+1} < a_N T \). The goal of \( \xi_j \) is to distinguish whether a mutation is faster than usual: we call a type \( j \) mutation an early type \( j \) mutation if it occurs in the time interval \([\tau_j, \xi_j]\). The fitness being relative to the mean, the earlier a mutant is, the stronger is its advantage to reproduce immediately after the mutation. The individual acquiring an early type \( j \) mutation, as well as its offsprings, are called early type \( j \) individuals. Schweinsberg showed that large families appear with the Bolthausen-Sznitman rates as a result of early mutations.

In [15, 16], \( \zeta \) denotes a stopping time up to which the evolution of the population can be controlled. Its definition requires a lot of technical considerations that are not relevant for our purposes, we therefore refer to Section 3.3 in [15] and equation (4.11) in [16] for equivalent definitions of \( \zeta \). In particular, for \( N \) large enough, it holds that \( \mathbb{P}(\zeta > a_N T) > 1 - \epsilon \). We shall work on \( \{\zeta > a_N T\} \) so that the properties below hold.

The next proposition gathers some properties of the model: point 1 is taken from Proposition 3.3 point 1 and point 3 of Proposition 3.6 in [15]. Then, points 2, 3, 5 are respectively drawn from point 2 and 3 of Proposition 3.3, Lemma 4.5 in [15], whereas the point 4 corresponds to points 1, 2, 3 in Proposition 4.4 in [16].

**Proposition 1.** For \( N \) large enough, the following hold:

1. For all \( j \geq k^* + 1 \) such that \( \tau_{j+1} \leq \zeta \wedge a_N T \), no early type \( j \) individual acquires a \( j + 1 \)-th mutation before time \( \tau_{j+1} \). Furthermore, it holds that

\[
\frac{a_N}{3k_N} \leq \tau_{j+1} - \tau_j \leq \frac{2a_N}{k_N},
\]

and on \( \{\zeta > a_N T\} \), we have \( \tau_{j+1} > a_N T \) for \( J := 3T k_N + k^* + 1 \), so the types greater than \( J \) have not appeared at time \( a_N T \) yet.

2. For all \( j \geq k^* + 1 \) and \( t \in [\tau_j + \frac{a_N}{4T k_N}, \tau_{j+1}] \cap [0, \zeta \wedge a_N T] \):

\[
(1 - 4\delta) e^{\int_{\tau_j}^t G_j(v) dv} \leq \tilde{W}_j(t) \leq (1 + 4\delta) e^{\int_{\tau_j}^t G_j(v) dv},
\]

where \( \tilde{W}_j \) denote the number of non-early type \( j \). Moreover, the upper bound holds for all \( t \in [\xi_j, \tau_{j+1}] \cap [0, \zeta \wedge a_N T] \).

3. For all \( j \geq k^* + 1 \) and \( t \in [\tau_{j+1}, \tau_{j+1} + a_N] \cap [0, \zeta \wedge a_N T] \):

\[
\left(1 - \frac{\delta}{\mu}\right) \frac{s}{\mu} e^{\int_{\tau_{j+1}}^t G_j(v) dv} \leq W_j(t) \leq \left(1 + \frac{\delta}{\mu}\right) \frac{s}{\mu} e^{\int_{\tau_{j+1}}^t G_j(v) dv}.
\]

4. For all \( j \geq k^* + 1 \) and \( t \in [\tau_j, \tau_{j+1}] \cap [0, \zeta \wedge a_N T] \):

\[
s(q_j - C_3) \leq G_j(t) \leq s(q_j + C_3),
\]

\[
sk_N (1 - 2\delta) \leq G_j(t) \leq sk_N (e + 2\delta),
\]

\[
k_N (1 - 2\delta) \leq q_j \leq k_N (e + 2\delta).
\]
5. For all $j \geq k^* + 1$ such that $\tau_{j+1} \leq \zeta \wedge a_N T$, we have
\[
\frac{s}{C_6 \mu} \leq e^{\int_{\tau_j}^{{\tau_{j+1}}} G_j(v)dv} \leq \frac{2s}{\mu},
\]
for some explicit constant $C_6$.

In [15], Proposition 3.6 point 1 shows that on $\{\zeta > a_N T\}$, $\tau_{k^* + 1} \leq 2a_N/k_N$ so that $\tau_{j(2)} > \tau_{k^* + 1}$. As explained in the introduction, it will be more convenient for us to study the process starting at time $\tau_{j(2)}$. We also note that by point 1 of Proposition [1] above and assumption (A3), on the event $\{\zeta > a_N T\}$, all the fitnesses of the individuals until time $a_N T$ are positive and therefore
\[
G_j(t) = s(j - M(t)) - \mu, \quad \forall t \leq a_N T \text{ and } j \leq J,
\]
where $J := 3T k_N + k^* + 1$ is from point 1 of the above proposition.

In [16], the study of the process backwards in time requires to consider only types $j$’s that belong to some set $I \subset N$, defined just before Lemma 6.2 in [16]. Its definition involves a fixed parameter $t_0 \in (T - 37, T - 2)$. Choosing $t_0 = T - 3$, one gets
\[
I = \{j_1, \ldots, j_2\}
\]
with
\[
\begin{align*}
    j_1 & := \max\{j : \tau_j^* \leq 2a_N\} - \lfloor 9\delta T k_N \rfloor, \\
    j_2 & := \max\{j : \tau_j^* \leq a_N(T - 1 + 19/k_N)\} + \lfloor 9\delta T k_N \rfloor,
\end{align*}
\]
where the $\tau_j^*$’s are some deterministic times, approximating the random $\tau_j$’s (see Equation (6.1) in [16]). The relevant informations for our purposes are given by Lemma 6.2 in [16], which shows that on the event $\{\zeta > a_N T\}$, it holds that $\tau_{j_1} < 2a_N$, and $j_2 \geq L + 9$, where $L$ is defined in Lemma 5.1 of [16] as
\[
L := \inf\{j : \tau_j \geq a_N(T - 1) - 3a_N/k_N\}.
\]

It entails that $\tau_{j_2} \geq \tau_L + 9a_N/3k_N$ by Proposition [1] point 1. Hence, $\tau_{j_2} > a_N(T - 1)$. We thus have that on the event $\{\zeta > a_N T\}$, $j(2) \in I$ and $j(T - 1) \in I$, so that we can use the results of Schweinsberg stated for $j \in I$, since then $2a_N \leq \tau_j \leq a_N(T - 1)$. In particular, we will often apply Proposition [1] choosing $j \in I$ on the event $\{\zeta > a_N T\}$, without recalling that this ensures $\tau_j \leq a_N T$.

We state a result on the time discretization that shall be useful later on.

**Lemma 1.** For all $j \in I$, conditionally given $\mathcal{F}_{\tau_j}^N$ and on the event $\{\zeta > \tau_{j+1}\}$, it holds that
\[
\frac{1 - 2\delta}{q_j} \leq \frac{\tau_{j+1} - \tau_j}{a_N} \leq \frac{1 + 2\delta}{q_j}.
\]

**Proof.** By Proposition [1] point 4, we know that $\sup_{t \in [\tau_j, \tau_{j+1}]} |G_j(t) - sq_j| \leq sC_3$. Equation (8.32) in [16] states that
\[
(1 - \delta)a_N \leq \int_{\tau_j}^{\tau_{j+1}} \frac{G_j(v)}{s}dv \leq (1 + \delta)a_N.
\]
Therefore we have that
\[
\frac{1 - \delta}{q_j + C_3} \leq \frac{\tau_{j+1} - \tau_j}{a_N} \leq \frac{1 + \delta}{q_j - C_3},
\]
which implies for \( N \) large enough that
\[
\frac{1 - 2\delta}{q_j} \leq \frac{\tau_{j+1} - \tau_j}{a_N} \leq \frac{1 + 2\delta}{q_j},
\]
since by Proposition \( \text{[11]} \) point 4, \( q_j \geq (1 - \delta)k_N \to \infty \) as \( N \to \infty \).

\[3.2 \quad \text{Fluctuations of the non-early } Y\text{-individuals}\]

Recall that we assigned \( y_N[s/\mu] \) individuals at time \( \tau_{j(2)} \) to group \( Y \), and \( (1 - y_N)[s/\mu] \) to group \( X \), with the competition explained just above Theorem \( \text{[1]} \). One sees when \( \alpha = 0 \) that
\[
\{W_j(t) : t \leq a_NT, j \geq 0\} = \{X_j(t) + Y_j(t) : t \leq a_NT, j \geq 0\}
\]
is exactly the model of Schweinsberg. We complete the filtration \( F^N \) to take into account the groups of the individuals. For all \( j \geq j(2) \), with an abuse of notation, we denote by \( Y_j(\tau_{j+1}-) \) the number of type \( j \) \( Y \)-individuals in the population immediately before the competition occurring at time \( \tau_{j+1} \) between groups \( X \) and \( Y \) (but such that \( W_j(\tau_{j+1}) = [s/\mu] \)).

The following lemma will allow us to control the fluctuations of the \( Y \)-individuals (and straightforwardly of the \( X \)-individuals).

\textbf{Lemma 2.} For all \( j \in I, j \geq j(2) \), the following process is a square integrable martingale:
\[
(Z_j^Y(t))_{t \in [\tau_{j+1}, \tau_{j+2}]} := Y_j(t)e^{-\int_{\tau_{j+1}}^t G_j(v)dv} - Y_j(\tau_{j+1}) - \int_{\tau_{j+1}}^t \mu Y_{j-1}(u)e^{-\int_{\tau_{j+1}}^u G_j(v)dv} du.
\]
Moreover, for \( N \) large enough, for all \( t \in [\tau_{j+1}, \tau_{j+2} \wedge \zeta] \), one has the following upper bound for its conditional variance:
\[
\text{Var}\left(Z_j^Y(t) \mid F^N_{\tau_{j+1}}\right) \leq \frac{21}{\mu k_N}.
\]

\textit{Proof.} The fact that \( Z_j^Y \) is a square integrable martingale has already been discussed above, with variance given by \( \text{[15]} \). The upper bound is then a direct consequence of Lemma 9.27 in \( \text{[15]} \).

\[3.3 \quad \text{Lemma 3.} \quad \text{For all } j \in I, j \geq j(2), \text{ it holds that}\]
\[
P\left(\mathbb{1}_{\{N > \tau_{j+1}\}} \sup_{t \in [\tau_{j}, \tau_{j+1}]} \left|Y_{j-1}(t) - Y_{j-1}(\tau_j)e^{\int_{\tau_j}^t G_j(v)dv} - \delta[s/\mu] \right| F^N_{\tau_j} \right) = o\left(\frac{1}{k_N}\right).
\]

\textit{Proof.} The statements of Proposition \( \text{[1]} \) hold, since we work on the event \( \{N > \tau_{j+1}\} \). Write
\[
Y_{j-1}(t) = e^{\int_{\tau_j}^t G_j(v)dv} \left(y[s/\mu] + \int_{\tau_j}^t \mu Y_{j-2}(u)e^{-\int_{\tau_j}^u G_j(v)dv} du + Z_{j-1}^Y(t)\right).
\]

We shall bound the two last terms in the above parentheses. By point 3 of Proposition \( \text{III} \) one has
\[
\mu Y_{j-2}(u) e^{-\int_{t_j}^u G_{j-1}(v) dv} \leq (1 + \delta) s e^{\int_{t_j}^u G_{j-1}(v) dv} e^{-\int_{t_j}^u G_{j-1}(v) dv},
\]
now remark that \( G_{j-1}(t) = G_{j-2}(t) + s \) and obtain
\[
= (1 + \delta) s e^{\int_{t_j}^\tau G_{j-1}(v) dv} e^{-s(u-\tau_{j-1})},
\]
then using the point 5 of Proposition \( \text{III} \)
\[
\leq (1 + \delta) \frac{2s^2}{\mu} e^{-s(u-\tau_{j-1})}.
\]
Hence,
\[
\int_{\tau_j}^t \mu Y_{j-2}(u) e^{-\int_{t_j}^u G_{j-1}(v) dv} du \leq (1 + \delta) \frac{2s}{\mu} (e^{-s(\tau_j - \tau_{j-1})} - e^{-s(t - \tau_{j-1})}).
\]
Point 1 in Proposition \( \text{III} \) gives \( s(\tau_j - \tau_{j-1}) \geq s \frac{\alpha}{3kN} = \log(s/\mu)/3kN \to \infty \) by assumption \( (A_2) \) such that \( \log(s/\mu)/3kN \log(kN) \to \infty \) as \( N \to \infty \). In particular, on \( \{ \zeta > \tau_{j+1} \} \), it holds that
\[
e^{-s(\tau_{j+1} - \tau_j)} = o\left( \frac{1}{kN} \right) \tag{19}
\]
We thus have shown that for \( N \) large enough,
\[
\int_{\tau_j}^t \mu Y_{j-2}(u) e^{-\int_{t_j}^u G_{j-1}(v) dv} du \leq 3s \frac{\mu}{\mu} \times O(1/kN).
\]
Furthermore, using Lemma 4 and the Doob’s maximal inequality for squared integrable martingales, one has
\[
P \left( I_{\{ \zeta > \tau_{j+1} \}} \sup_{t \in [\tau_j, \tau_{j+1}]} \left| Z_{j-1}(t) \right| > \frac{s^2}{\mu} \left| F_{\tau_j} \right| \right) \leq \frac{\mu^2}{s^2} \sup_{t \in [\tau_j, \tau_{j+1}]} \text{Var} \left( I_{\{ \zeta > \tau_{j+1} \}} Z_{j-1}(t) | F_{\tau_j} \right) \leq \frac{21\mu}{s^4kN(1 - \epsilon)} = o \left( \frac{1}{kN} \right),
\]
which completes the proof. \( \blacksquare \)

We state a result that directly follows from Lemma 9.12 in [15] (and therefore whose proof is omitted):

**Lemma 4.** For \( j \in I, j \geq j(2), \) let \( \bar{W}_j \) be the process which counts the number of non-early individuals, i.e. that obtain a \( j \)th mutation during \( [\xi_j, \tau_{j+1}] \) and their descendants of type \( j \), and \( \bar{Z}_j \) the associated martingale, that is for all \( t \in [\xi_j, \tau_{j+1}] \):
\[
\bar{Z}_j(t) := e^{-\int_{\xi_j}^t G_{j}(v) dv} \bar{W}_j(t) - \int_{\xi_j}^t \mu W_{j-1}(u) e^{-\int_{\xi_j}^u G_{j}(v) dv} du.
\]
Then, its conditional variance at time \( \tau_{j+1}^- \) satisfies
\[
\text{Var} \left( I_{\{ \zeta > \tau_{j+1} \}} \bar{Z}_j(\tau_{j+1}^-) | F_{\xi_j} \right) \leq \frac{Ce^{\int_{\xi_j}^{\tau_j} G_{j}(v) dv}}{skN^2}.
\]
Moreover, the same upper bound holds for both martingales \( \bar{Z}_j^X \) and \( \bar{Z}_j^Y \) (with obvious notations).
3.3 Proportions immediately after an early mutation

We now investigate the impact of an early mutation on the proportions. The next lemma essentially states that the proportion of $Y$-individuals from $\tau_j$ to $\tau_{j+1}$ is mostly determined by the number of early type $j+1$ individuals.

Lemma 5. Let $j \in I$, $j \geq j(2)$ and let $S$ be the proportion of early type $j$ individuals at time $\tau_{j+1}$. Conditionally given $F_{\tau_j}$, on the event $\{\zeta > \tau_{j+1}\}$, the following holds with probability at least $1 - C\epsilon/k_N$: If $S > 0$, then the early individuals have the same ancestor at time $\tau_j$, and if it belongs to the group $X$, then

$$Y_{j-1}(\tau_j)(1 - S) - 8\delta[s/\mu] \leq Y_j(\tau_{j+1}) \leq Y_{j-1}(\tau_j)(1 - S) + 8\delta[s/\mu].$$

Similarly, if the ancestor belongs to the group $Y$, then

$$S[s/\mu] + Y_{j-1}(\tau_j)(1 - S) - 8\delta[s/\mu] \leq Y_j(\tau_{j+1}) \leq S[s/\mu] + Y_{j-1}(\tau_j)(1 - S) + 8\delta[s/\mu].$$

Proof. Recall the notation $\tilde{W}_j$ and $\tilde{Y}_j$ for the processes following the non-early type $j$ individuals, respectively the non-early type $j$ $Y$-individuals. We write

$$\frac{\tilde{Y}_j(\tau_{j+1})}{W_j(\tau_{j+1})} = \frac{e^{\int_{\xi_j}^{\tau_{j+1}} G_j(v)dv} \left( \int_{\xi_j}^{\tau_{j+1}} \mu Y_{j-1}(u)e^{-\int_{\xi_j}^{u} G_j(v)dv} du + \tilde{Z}_j(\tau_{j+1}) \right)}{W_j(\tau_{j+1})}.$$

Using Proposition [1] point 2 and Lemma [3], one has with probability at least $1 - o(1/k_N)$ that

$$\frac{\tilde{Y}_j(\tau_{j+1})}{W_j(\tau_{j+1})} \leq \frac{e^{-\int_{\xi_j}^{\tau_{j+1}} G_j(v)dv} \left( Y_{j-1}(\tau_j) \right)}{1 - 4\delta} \left( \int_{\xi_j}^{\tau_{j+1}} se^{-s(u-\xi_j)} du e^{\int_{\xi_j}^{u} G_j(v)dv} \right) \left( \int_{\xi_j}^{\tau_{j+1}} e^{-\int_{\xi_j}^{u} G_j(v)dv} du + \tilde{Z}_j(\tau_{j+1}) \right). \tag{20}$$

We compute $\int_{\xi_j}^{\tau_{j+1}} se^{-s(u-\xi_j)} du = 1 - e^{-s(\tau_{j+1}-\xi_j)}$, and claim that this converges to 1 as $N \to \infty$. Indeed, we first note that, on the event $\{\zeta > \tau_{j+1}\}$, point 4 of Proposition [1] entails that

$$e^{-s(\xi_j-\tau_j)} = (s\xi_j)^{1/q_j} e^{-b/q_j} \to 1. \tag{21}$$

We then write

$$e^{-s(\tau_{j+1}-\xi_j)} = e^{-s(\tau_{j+1}-\tau_j)} e^{s(\xi_j-\tau_j)} = o(1/k_N),$$

where we have used [19]. By hypothesis $(A_2)$, we see that $s\xi_j N \to \infty$ as $N \to \infty$. We also have that

$$e^{-\int_{\xi_j}^{\tau_{j+1}} G_j(v)dv} e^{\int_{\xi_j}^{\tau_{j+1}} G_j(v)dv} = e^{-s(\xi_j-\tau_j)} \to 1.$$
By Proposition 1 point 4, we see that
\[ s\delta \int_{\xi_j}^{\tau_j+1} e^{-f_{ij}^\nu G_j(v)dv} du \leq s\delta \int_{\xi_j}^{\tau_j+1} e^{-s(q_j-C_3)(u-\xi_j)} du \leq \frac{\delta}{q_j-C_3} \leq \delta, \]
for \( N \) large enough. Coming back to (20), this gives
\[ \frac{\tilde{Y}_j(\tau_j+1)}{W_j(\tau_j+1)} \leq \frac{1}{1-4\delta} \left( \frac{Y_{j-1}(\tau_j)}{|s/\mu|} + e^{f_{ij}^\nu G_j(v)dv} \left( \delta + \tilde{Z}_j^Y(\tau_j+1) \right) \right). \]  
(22)

Applying Lemma 4 and Doob’s maximal inequality for squared integrable martingales, one has
\[ \mathbb{P}\left( \mathbb{I}_{\{\zeta > \tau_{j+1}\}} \left| \frac{\tilde{Z}_j^Y(\tau_{j+1}-)}{W_j(\tau_{j+1}-)} > \delta e^{f_{ij}^\nu G_j(v)dv} \right| \mathcal{F}_{\xi_j} \right) \leq C e^{-f_{ij}^\nu G_j(v)dv} \frac{\delta}{\delta^2k_N^2}. \]

On the event \( \{\zeta > \tau_{j+1}\} \) a double application of point 4 of Proposition 1 gives
\[ e^{-f_{ij}^\nu G_j(v)dv} \leq e^{-s(q_j-C_3)(\zeta_j-\tau_j)} \leq sq_j e^{-b}(1+\delta) \leq Csk_N e^{-b}. \]

(The constant \( C_3 \) has been absorbed in \( C \), which does not depend on the parameters \( \epsilon, \delta, T \).)

Hence,
\[ \mathbb{P}\left( \mathbb{I}_{\{\zeta > \tau_{j+1}\}} \left| \frac{\tilde{Z}_j^Y(\tau_{j+1}-)}{W_j(\tau_{j+1}-)} > \delta e^{f_{ij}^\nu G_j(v)dv} \right| \mathcal{F}_{\xi_j} \right) \leq C e^{-b} \frac{\delta}{\delta^2k_N} \leq C k_N, \]
where we used (8) and (16) for the last inequality. This result combined with (22) entails that, on the event \( \{\zeta > \tau_{j+1}\} \), with a probability greater than \( 1-C/\delta N \):
\[ \frac{\tilde{Y}_j(\tau_{j+1}-)}{W_j(\tau_{j+1}-)} \leq \frac{1}{1-4\delta} \left( \frac{Y_{j-1}(\tau_j)}{|s/\mu|} + 2\delta \right) \leq \frac{Y_{j-1}(\tau_j)}{|s/\mu|} + 7\delta. \]

The same reasoning also gives
\[ \frac{\tilde{Y}_j(\tau_{j+1}-)}{W_j(\tau_{j+1}-)} \geq \frac{1}{1+4\delta} \left( \frac{Y_{j-1}(\tau_j)}{|s/\mu|} - 3\delta \right) \geq \frac{Y_{j-1}(\tau_j)}{|s/\mu|} - 8\delta. \]

Since \( \tilde{W}(\tau_{j+1}-) = (1-S)|s/\mu| \), homogenizing the bounds, we get
\[ \frac{Y_{j-1}(\tau_j)}{|s/\mu|} - 8\delta \leq \frac{\tilde{Y}_j(\tau_{j+1}-)}{W_j(\tau_{j+1}-)} \leq \frac{Y_{j-1}(\tau_j)}{|s/\mu|} + 8\delta. \]

To conclude, conditionally given \( \mathcal{F}_{\tau_j}^N \), Lemma 7.5 in (10) bounds from above the probability that two early mutations survive by \( 2e^{2b}/q_j^2 \leq 3e^{2b}/k_N^2 \). Then, excluding this event, if there is an early mutation in the group \( X \), easy calculations lead to
\[ Y_{j-1}(\tau_j)(1-S) - 8\delta|s/\mu| \leq Y_j(\tau_{j+1}-) \leq Y_{j-1}(\tau_j)(1-S) + 8\delta|s/\mu|, \]
\[ S|s/\mu| + X_{j-1}(\tau_j)(1-S) - 8\delta|s/\mu| \leq X_j(\tau_{j+1}-) \leq S|s/\mu| + X_{j-1}(\tau_j)(1-S) + 8\delta|s/\mu|. \]

The cases where the early mutant is a \( Y \)-individual is identical, which concludes the proof. ■

13
To ensure that our description of the evolution of the proportions is accurate enough, we introduce a stopped discrete time process as follows. For all $j \in I$, $j \geq j(2) - 1$, we define

$$Y^N_j := \frac{Y_j \wedge (j(\zeta) - 1)(\tau_{j+1} \wedge j(\zeta))}{[s/\mu]},$$

Which follows the proportion stopped at the last type $j$ before $j(\zeta)$. We denote by $Y^N_{j-}$ the proportion of type $j \wedge (j(\zeta) - 1)$ individuals at time $\tau_{j+1} \wedge j(\zeta)$ right before the killing described by (7). We stress that the event $\{\zeta > \tau_{j+1}\}$ is included in the above definition and this fact will be kept implicit when working with $Y$.

Let $p_S$ be the distribution of $S$ as in Lemma 5 above. We now give a lemma controlling the two first moments of the proportions’ increments, when there is no early mutation, or one that does not generate a too large family.

**Lemma 6.** Let $S$ be the proportion of early type $j$ individuals at time $\tau_{j+1}$ (potentially, $S$ can be 0). For all $j \in I$, $j \geq j(2)$, it holds that

$$E \left( Y^N_{j-} - Y^N_{j-1} \mid \mathcal{F}^N_{\tau_j} \right) = o(1/k_N),$$

$$E \left( \mathbb{1}_{\{S \leq \epsilon\}} (Y^N_{j-} - Y^N_{j-1}) \mid \mathcal{F}^N_{\tau_j} \right) = o(1/k_N),$$

$$E \left( \mathbb{1}_{\{S \leq \epsilon\}} (Y^N_{j-} - Y^N_{j-1})^2 \mid \mathcal{F}^N_{\tau_j} \right) \leq \frac{C\epsilon}{k_N}.$$

Moreover, the $o(1/k_N)$ in the first equation is independent from the parameters $\delta, \epsilon, T$ of the model.

**Proof.** Fix $j \in I$, $j \geq j(2)$. We first note that

$$Y^N_{j-} - Y^N_{j-1} = \mathbb{1}_{\{\zeta > \tau_{j+1}\}} \frac{Y_j(\tau_{j+1} - \tau_j)}{[s/\mu]}.$$

We call a $Y$-individual of type $j$ at time $\tau_{j+1}$ - **good** if his ancestor at time $\tau_j$ is of type $j - 1$. We denote by $\tilde{Y}_j(\tau_{j+1} -)$ the number of good $Y$ individuals at time $\tau_{j+1}$, and $K_j$, respectively $K_{Y,j}$ the number of type $j$ individuals in the population, respectively in group $Y$, at time $\tau_{j+1}$ that are not good. We have

$$Y_j(\tau_{j+1} -) = K_{Y,j} + \tilde{Y}_j(\tau_{j+1} -).$$

We pick an individual uniformly at random among the $[s/\mu]$ individuals of type $j$ at time $\tau_{j+1}$. Note that he belongs to the group of good $Y$-individuals if and only if his ancestor at time $\tau_j$ is in the group $Y$ with type $j - 1$, and we call this event $B$. Let $j_{\text{anc}} \in \mathbb{N}$ be the type of his ancestor at time $\tau_j$. We have in particular that $P(B \mid \mathcal{F}^N_{\tau_j}, j_{\text{anc}} = j - 1, \zeta > \tau_{j+1}, S \leq \epsilon) = Y_{j-1}(\tau_j)/[s/\mu]$, since given this conditioning, his ancestor is independent from $S$ and chosen uniformly at random among the $[s/\mu]$ individuals of type $j - 1$ at time $\tau_j$. Using that
\{ζ > τ_{j+1}, S ≤ ε\} is \mathcal{F}_{t_{j+1}}\text{-measurable, we then write}

\[
\mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}, S ≤ ε\}} \frac{\hat{Y}_j(τ_{j+1}-)}{[s/μ]} \bigg| \mathcal{F}_{t_j}^N \right)
= \mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}, S ≤ ε\}} \mathbb{P}\left( B \big| \mathcal{F}_{t_{j+1}}^N \right) \bigg| \mathcal{F}_{t_j}^N \right)
= \mathbb{P}\left( B \big| \mathcal{F}_{t_j}^N \right) \mathbb{P}\left( j_{\text{anc}} = j-1, ζ > τ_{j+1}, S ≤ ε \right)
= \frac{Y_{j-1}(τ_j)}{[s/μ]} \mathbb{P}\left( j_{\text{anc}} = j-1, ζ > τ_{j+1}, S ≤ ε \right).
\]

Basic properties of probability measures entail that

\[
\frac{Y_{j-1}(τ_j)}{[s/μ]} \mathbb{P}(ζ > τ_{j+1}, S ≤ ε | \mathcal{F}_{t_j}^N) \leq \mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}, S ≤ ε\}} \frac{\hat{Y}_j(τ_{j+1}-)}{[s/μ]} \bigg| \mathcal{F}_{t_j}^N \right) \leq \frac{Y_{j-1}(τ_j)}{[s/μ]} \mathbb{P}(ζ > τ_{j+1}, S ≤ ε | \mathcal{F}_{t_j}^N).
\]

Since \( \mathbb{P}(ζ > τ_{j+1}, j_{\text{anc}} ≠ j-1 | \mathcal{F}_{t_j}^N) = \mathbb{E}(\mathbb{I}_{\{ζ > τ_{j+1}\}} K_j / [s/μ] | \mathcal{F}_{t_j}^N) \), we have shown that

\[
\mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}\}} (Y_j - Y_{j-1}) \bigg| \mathcal{F}_{t_j}^N \right) \leq \mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}\}} \frac{K_{Y,j}}{[s/μ]} \bigg| \mathcal{F}_{t_j}^N \right) + \mathbb{P}(ζ > τ_{j+1}, j_{\text{anc}} ≠ j-1 | \mathcal{F}_{t_j}^N)
\leq 2 \mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}\}} \frac{K_j}{[s/μ]} \bigg| \mathcal{F}_{t_j}^N \right) ≤ \left( \frac{μ}{s} \right)^{1/3κ_N}, \tag{23}
\]

where the last inequality is from [15] Lemma 6.3 and, taking the logarithm and using assumption \( A_2 \), one can show that the bound is \( o(k_N) \). The proof for the same bound without \( \mathbb{I}_{\{S ≤ ε\}} \) is identical, as this event played no particular role in the above computations.

We turn our attention to the moment of order two. Suppose now that we independently sample two individuals, possibly the same, uniformly at random among the \([s/μ]\) individuals of type \( j \) at time \( τ_{j+1} \). Denote \( j_{\text{anc}} \) and \( j'_{\text{anc}} \) the types of their respective ancestors at time \( τ_j \) and let \( B' \) be the event that they both belong to the good \( Y \) group. Let \( D \) be the event that the two ancestors are different with \( j_{\text{anc}} = j'_{\text{anc}} = j-1 \). In particular, given \( D \), the ancestor of the first individual is chosen uniformly at random among the \([s/μ]\) individuals of type \( j-1 \) at time \( τ_j \), and then the ancestor of the second one is chosen uniformly at random among the \([s/μ] - 1 \) that remain, the two ancestors being independent from \( S \). We get that

\[
\mathbb{E}\left( \mathbb{I}_{\{ζ > τ_{j+1}, S ≤ ε\}} \left( \frac{\hat{Y}_j(τ_{j+1}-)}{[s/μ]} \right)^2 \bigg| \mathcal{F}_{t_j}^N \right) = \mathbb{P}(B' \cap \{ζ > τ_{j+1}, S ≤ ε\} | \mathcal{F}_{t_j}^N)
= \mathbb{P}(B' | \mathcal{F}_{t_j}^N, D \cap \{ζ > τ_{j+1}, S ≤ ε\}) \mathbb{P}(D \cap \{ζ > τ_{j+1}, S ≤ ε\} | \mathcal{F}_{t_j}^N)
+ \mathbb{P}(B' \cap D^c \cap \{ζ > τ_{j+1}, S ≤ ε\} | \mathcal{F}_{t_j}^N)
\leq \left( \frac{Y_{j-1}(τ_j)}{[s/μ]} \right)^2 \mathbb{P}(ζ > τ_{j+1}, S ≤ ε | \mathcal{F}_{t_j}^N) + \mathbb{P}(B' \cap D^c \cap \{ζ > τ_{j+1}, S ≤ ε\} | \mathcal{F}_{t_j}^N).
\]
Note that $B' \cap D^c$ is included in the event that the two sampled individuals have the same ancestor of type $j - 1$ at time $\tau_j$. The probability to pick twice the same individual is $1/[s/\mu]$. The probability that two different individuals have the same ancestor on the event $\{S \leq \epsilon\}$ is bounded in [10]. More precisely, Lemmas 6.5, 6.6 and Equation (8.16) in Lemma 8.8 together show that $\mathbb{P}(B' \cap D^c \cap \{\zeta > \tau_{j+1}, S \leq \epsilon\}|\mathcal{F}_{\tau_j}^N) \leq C\epsilon/k_N$. We obtain that

$$
\mathbb{E} \left( 1_{\{S \leq \epsilon\}} (Y_{j-1}^N - Y_{j-1}^N)^2 \middle| \mathcal{F}_{\tau_j}^N \right)
\leq 2\mathbb{E} \left( 1_{\{\zeta > \tau_{j+1}\}} \left( \frac{K_j}{s/\mu} \right)^2 \middle| \mathcal{F}_{\tau_j}^N \right)
$$

$$
+ 2\mathbb{E} \left( 1_{\{\zeta > \tau_{j+1}, S \leq \epsilon\}} \frac{1}{s/\mu}^2 \left( \hat{Y}_j(\tau_{j+1} - )^2 - 2\hat{Y}_j(\tau_{j+1} - )Y_{j-1}(\tau_j) + Y_{j-1}(\tau_j)^2 \right) \middle| \mathcal{F}_{\tau_j}^N \right)
$$

$$
\leq C\epsilon/k_N,
$$
as claimed. ■

We conclude this last subsection of the toolbox with a reformulation of the result of Schweinsberg in [10] showing that the law of $S$ can be well approximated, on $\{S > \epsilon\}$, by the rates corresponding to Bolthausen-Sznitman coalescent.

**Lemma 7.** For $N$ large enough, for all $j \in I$, $j \geq j(2)$, conditionally given $\mathcal{F}_{\tau_j}^N$ and on the event $\{\zeta > \tau_{j+1}\}$, for any $g \in C^\infty([0,1])$, it holds that

$$
\left| q_j \int_{(\epsilon,1)} g(x)p_S(dx) - \int_{\epsilon}^1 g(x)\frac{dx}{x^2} \right| \leq C(\|g\|_\infty + \|g'\|_\infty)\epsilon,
$$

where $S$ is the number of early type $j$ individuals at time $\tau_{j+1}$.

**Proof.** Let $\nu(dx) = dx/x^2$, $x \in (0,1]$. Lemma 7.8 in [10] shows that for all $y \in (\epsilon,1-\delta]$, it holds\(^\dagger\) that

$$
|q_jp_S((y,1)) - \nu((y,1))| \leq 14\delta\nu((y,1)) \leq 14\frac{\delta}{\epsilon}.
$$

We write

$$
\left| q_j \int_{(\epsilon,1)} g(x)p_S(dx) - \int_{\epsilon}^1 g(x)\nu(dx) \right| = \left| \int_{\epsilon}^1 dyq_j(y)p_S - \nu)((y,1) - g(\epsilon)(q_jp_S - \nu)((\epsilon,1]) \right|
$$

$$
\leq 14\frac{\delta}{\epsilon}\|g'\|_\infty + \|g'\|_\infty \int_{1-\delta}^1 dyq_j(y)p_S + \nu((1-\delta,1)) + 14\frac{\delta}{\epsilon}\|g\|_\infty \leq C\frac{\delta}{\epsilon}(\|g\|_\infty + \|g'\|_\infty).
$$

We conclude using that $\delta/\epsilon < \epsilon$ by (8). ■

\(^\dagger\)We implicitly use that the event in [10] equation (7.48) has probability going to 1 as $N \to \infty$, see Lemmas 7.4 and 7.7 of the same paper. Roughly speaking, on this event the early mutants are coupled with a branching process introduced in Section 7.2 of the same paper, allowing to approximate the law of $S$. 

16
4 Convergence towards the SDE.

We first establish the tightness of \((Y^N_t)_{t \in [2, T-1]}\). Recall that at each step from \(\tau_j\) to \(\tau_{j+1}\), the killing takes place, as described in \(\text{(7)}\).

**Lemma 8.** Recall that \(Y^N_2 = y_N\) with \(y_N \to y \in (0, 1)\). The sequence \(\{(Y^N(t))_{t \in [2, T-1]}; N \in \mathbb{N}\}\) is tight.

**Proof.** The proof uses Aldous’ criterion for tightness, stated e.g. in [10] Chapter VI Theorem 4.5. Let \(\lambda, \theta > 0\) and let \(\sigma, \sigma'\) denote any two stopping times with respect to the filtration \(\mathcal{F}_N\), that are bounded by \(T - 1\), and such that \(\sigma \leq \sigma' \leq \sigma + \theta\). Splitting the following probability

\[
\mathbb{P}\left(|Y^N_{\sigma'} - Y^N_{\sigma}| > \lambda\right) \leq \mathbb{P}\left(1_{\{\zeta > aNT\}}|Y^N_{\sigma'} - Y^N_{\sigma}| > \lambda\right) + \epsilon
\]

\[
\leq \lambda^{-2} \mathbb{E}\left(\mathbb{E}\left(1_{\{\zeta > aNT\}} \left(\frac{Y^N_{\sigma} - Y^N_{\sigma'}}{\mathcal{F}_{\tau_j(\sigma)}}\right)^2\right)\right) + \epsilon.
\]

We rewrite the conditional expectation as

\[
\mathbb{E}\left(1_{\{\zeta > aNT\}}\left(\sum_{j=\sigma}^{(j+1)} \frac{Y^N_j - Y^N_{j-1}}{\mathcal{F}_{\tau_j(\sigma)}}\right)^2\right)
\]

Recall that on the event \(\{\zeta > aNT\}\), the number of \(j \geq 1\) such that \(\sigma \leq \tau_j \leq \sigma'\) is at most \(3kN\theta\) by Proposition 1 point 1. We bound the above by

\[
\mathbb{E}\left(\sum_{j=\sigma}^{(j+1)} (Y^N_j - Y^N_{j-1})^2\right)\mathcal{F}_{\tau_j(\sigma)}
\]

\[
\leq \mathbb{E}\left(\sum_{j=\sigma}^{(j+1)} (Y^N_j - Y^N_{j-1})^2\right)\mathcal{F}_{\tau_j(\sigma)}
\]

We now bound the first sum of the right-hand side. Let \(S_j\) be the number of early type \(j\) individuals at time \(\tau_{j+1}\). Recall \(\text{(7)}\), bounding the number of individuals killed by \(\alpha \Delta_j\), we see that

\[
\sum_{j=\sigma}^{(j+1)} \mathbb{E}\left(\mathbb{E}\left((Y^N_j - Y^N_{j-1})^2\right)\mathcal{F}_{\tau_j(\sigma)}\right)
\]

\[
\leq 2 \sum_{j=\sigma}^{(j+1)} \mathbb{E}\left(\mathbb{E}\left(\alpha^2 \Delta_j^2 + (Y^N_j - Y^N_{j-1})^2\right)\mathcal{F}_{\tau_j(\sigma)}\right)
\]

Using our Proposition 1 point 1 to bound \(\Delta_j\), our Lemma 6 and Lemma 7.8 in [16] for the other term, we can bound the right-hand side by

\[
2 \sum_{j=\sigma}^{(j+1)} \left(\frac{\alpha^2 C}{k_N} + \frac{C\epsilon}{k_N} + \mathbb{E}\left(\mathbb{P}\left(S_j > \epsilon, \zeta > \tau_{j+1}\right)\mathcal{F}_{\tau_j(\sigma)}\right)\right) \leq C\epsilon + \theta \epsilon \leq C\frac{\theta}{\epsilon}.
\]
We next turn our attention to the double sum, we write
\[
\sum_{j<\ell}^{j(\sigma)+n} (Y^N_j - Y^N_{j-1}) (Y^N_{\ell} - Y^N_{\ell-1}) \leq \sup_{n \leq [3k_N\theta]} \sum_{j,\ell=j(\sigma)}^{j(\sigma)+n} (Y^N_j - Y^N_{j-1}) (Y^N_{\ell} - Y^N_{\ell-1}),
\]
therefore we have that
\[
\left| \mathbb{E} \left( \sum_{j,\ell=j(\sigma)}^{j(\sigma)-1} (Y^N_j - Y^N_{j-1}) (Y^N_{\ell} - Y^N_{\ell-1}) \mid \mathcal{F}^N_{\tau_j(\sigma)} \right) \right|
\]
\[
\leq \sup_{n \leq [3k_N\theta]} \left| \sum_{j,\ell=j(\sigma)}^{j(\sigma)+n} \mathbb{E} \left( (Y^N_j - Y^N_{j-1}) \mathbb{E} \left( Y^N_{\ell} - Y^N_{\ell-1} \mid \mathcal{F}^N_{\tau_j(\sigma)} \right) \right) \right|
\]
\[
\leq \left( \left( \frac{\mu}{s} \right)^{1/3k_N} + \alpha \Delta_j \right)^{j(\sigma)+[3k_N\theta]} \sum_{j,\ell=j(\sigma)}^{j(\sigma)+[3k_N\theta]} \mathbb{E} \left( |Y^N_j - Y^N_{j-1}| \mid \mathcal{F}^N_{\tau_j(\sigma)} \right),
\]
where we used (23) and the expression for the killing (17) for the last bound (recall that (23) remains valid without \( I_{\{S \leq \sigma\}} \)). Clearly, we can bound the above by
\[
\left( \frac{\mu}{s} \right)^{1/3k_N} [3k_N\theta] + \frac{C}{k_N} [3k_N\theta],
\]
where \( C \) is independent from \( N, \epsilon, \delta \) and \( \theta \). Taking the logarithm, one obtains an upper bound for the first term of the same order as \( \log(k_N)(1 - \log(s/\mu)/3k_N \log(k_N)) \), which goes to \(-\infty\) as \( N \to \infty \), by assumption \( A_2 \). Hence, coming back to (25), we have shown that for all \( \lambda > 0 \), it holds that
\[
\lim \limsup_{\theta \to 0} \sup_{N \to \infty} \mathbb{P} (|Y^N_\sigma - Y^N_0| > \lambda) \leq \lambda^{-2} \lim_{\theta \to 0} \left( \lim_{N \to \infty} \left( \frac{\mu}{s} \right)^{1/3k_N} [3k_N\theta] + \frac{C}{\epsilon} \theta + \epsilon \right) = \lambda^{-2} \epsilon.
\]
Since the left-hand side does not depend on \( \epsilon \), its value is simply 0, which shows that the sequence is tight by Aldous’ criterion for tightness.

Even though \( \{Y^N_j; j \in I, j \geq j(2) - 1\} \) is not Markovian, we shall mimic a classical method for showing convergence of a Markov process through its infinitesimal generator.

**Lemma 9.** On the event \( \{\zeta > \tau_{j+1}\} \), for all \( f \in C^\infty([0,1]) \), all \( j \in I, j \geq j(2) \) and \( N \) large enough, it holds that
\[
\left| \Delta_j \mathbb{E} \left( f(Y^N_j) - f(Y^N_{j-1}) \mid \mathcal{F}^N_{\tau_j} \right) + \alpha Y^N_{j-1}(1 - Y^N_{j-1}) f'(Y^N_{j-1}) - \int_0^1 \frac{dx}{x^2} \int_0^1 du \left( f(Y^N_{j-1}(1-x)) \mathbb{1}_{u \leq Y^N_{j-1}} - f(Y^N_{j-1}) \right) \right| \leq C \epsilon (||f||_\infty + ||f'||_\infty + ||f''||_\infty).
\]
Proof. We write
\[ E \left( f(Y_j^N) - f(Y_{j-1}^N) \mid \mathcal{F}_{\tau_j}^N \right) = E \left( \mathbf{1}_{\{\xi > \tau_{j+1}\}} \left( f(Y_j^N - \alpha \Delta_j Y_j^N (1 - Y_j^N)) - f(Y_{j-1}^N) \right) \mid \mathcal{F}_{\tau_j}^N \right) \]
\[ = E \left( \mathbf{1}_{\{\xi > \tau_{j+1}\}} \left( f(Y_j^N - \alpha \Delta_j Y_j^N (1 - Y_j^N)) - f(Y_{j-1}^N) + f(Y_j^N) - f(Y_{j-1}^N) \right) \mid \mathcal{F}_{\tau_j}^N \right) \]
\[ = E \left( \mathbf{1}_{\{\xi > \tau_{j+1}\}} \left( - \alpha \Delta_j Y_j^N (1 - Y_j^N) f'(Y_j^N) + O(\Delta_j^2) ||f''||_\infty + f(Y_j^N) - f(Y_{j-1}^N) \right) \mid \mathcal{F}_{\tau_j}^N \right), \]
where we used the Taylor-Lagrange formula. Note that in the above expression, \( \mathbf{1}_{\{\xi > \tau_{j+1}\}} |O(\Delta_j^2)| \leq 1/q_j^2 \), thanks to Lemma 1. We focus on the first term. Let \( S \) be the proportion of early type \( j \) individuals at time \( \tau_{j+1} \). By Lemma 5, we have that
\[ \left| E \left( \mathbf{1}_{\{\xi > \tau_{j+1}, S = 0\}} \alpha \Delta_j Y_j^N (1 - Y_j^N) f'(Y_j^N) \mid \mathcal{F}_{\tau_j}^N \right) - \alpha \Delta_j Y_{j-1}^N (1 - Y_{j-1}^N) f'(Y_{j-1}^N) \right| \]
\[ \leq E \left( \mathbf{1}_{\{\xi > \tau_{j+1}\}} \alpha \Delta_j \left| f'(Y_j^N) \right| \left( \frac{C \epsilon}{k_N} ||f'||_\infty + 8\delta(2||f'||_\infty + ||f''||_\infty) \right) \right) \leq \frac{C \delta}{q_j} (||f'||_\infty + ||f''||_\infty), \]
thanks to Lemma 1. Moreover, Lemma 7.5 in [16] entails that \( \mathbb{P}(\xi > \tau_{j+1}, S > 0 | \mathcal{F}_{\tau_j}^N) \leq (1 + 2\delta)e^{\beta}/q_j \). Therefore, again by Lemma 1, we have
\[ \left| E \left( \mathbf{1}_{\{\xi > \tau_{j+1}, S \geq 0\}} \alpha \Delta_j Y_j^N (1 - Y_j^N) f'(Y_j^N) \mid \mathcal{F}_{\tau_j}^N \right) \right| \leq \frac{C e^{\beta} ||f'||_\infty}{q_j^2}. \]

Next, we use the Taylor-Lagrange formula, which ensures the existence of \( \xi \) strictly between \( Y_{j-1}(\tau_j) \) and \( Y_j(\tau_j + 1)/[s/\mu] \) such that
\[ \mathbf{1}_{\{\xi > \tau_{j+1}, S \leq \epsilon\}} \left( f(Y_j^N) - f(Y_{j-1}^N) \right) \]
\[ = \mathbf{1}_{\{\xi > \tau_{j+1}, S \leq \epsilon\}} \left( (Y_j^N - Y_{j-1}^N) f'(Y_j^N - Y_{j-1}^N) + (Y_j^N - Y_{j-1}^N)^2 \frac{f''(\xi)}{2} \right). \]
Thanks to Lemma 6 and Proposition 1, point 4, we see that
\[ \left| E \left( \mathbf{1}_{\{\xi > \tau_{j+1}, S \leq \epsilon\}} \left( f(Y_j^N) - f(Y_{j-1}^N) \right) \mid \mathcal{F}_{\tau_j}^N \right) \right| \leq \frac{C \epsilon}{q_j} (||f'||_\infty + ||f''||_\infty) \]

We now turn our attention on the difference when an early mutation generates a large family. Let \( p_S \) denote the conditional distribution of \( S \) given \( \mathcal{F}_{\tau_j} \), supported on \{0, 1/[s/\mu], \ldots, 1\}. Note that if an early mutation occurs as described in Lemma 5 on the event \( \{\tau_{j+1} < \xi\} \) and given \( \mathcal{F}_{\tau_j}^N \), the individual who generates the large family is chosen uniformly at random among the \([s/\mu]\) type \( j \) individuals at time \( \tau_j \). Hence, the conditional probability that the early individual is in group \( Y \), respectively \( X \), is \( Y_j^N \), respectively \( 1 - Y_j^N \). Thanks to Lemma 5, we can write
\[ E \left( \mathbf{1}_{\{\xi > \tau_{j+1}, S > \epsilon\}} \left( f(Y_j^N) - f(Y_{j-1}^N) \right) \mid \mathcal{F}_{\tau_j}^N \right) \]
\[ = E + \int_{(\epsilon, 1]} p_S(dx) \left( Y_j^N f(Y_j^N(1 - x) + x) + (1 - Y_j^N)f(Y_j^N(1 - x)) - f(Y_{j-1}^N) \right), \]
where $E$ is the error coming from the approximation in Lemma 5 and the probability that this approximation does not hold. In particular, we have that

$$|E| \leq p_S((\epsilon, 1]) \left( \sup_{-\delta < z < \delta} |f(Y_{j-1}^N(1-x) + x + z) - f(Y_{j-1}^N(1-x) + x)| + \sup_{-\delta < z < \delta} |f(Y_{j-1}^N(1-x) + z) - f(Y_{j-1}^N(1-x))| + \frac{C_\epsilon}{k_N} \|f\|_\infty \right)$$

$$\leq \frac{1 + 13\delta}{\epsilon q_j} 16\delta \|f'\|_\infty + \frac{C_\epsilon}{q_j} \|f\|_\infty \leq C_\epsilon \|f\|_\infty + \|f'\|_\infty,$$

where we used Lemma 7.8 in [16], Proposition 1 point 4, and that $\epsilon > \delta^3$ by (8). Lemma 7 allows us to write

$$\left| \mathbb{E} \left( I_{\{z > \tau_{j+1}, S > \ell\}} (f(Y_{j-1}^N) - f(Y_{j-1}^N)) \bigg| \mathcal{F}_{\tau_j}^N \right) \right|$$

$$- \frac{1}{q_j} \int_0^1 \frac{dx}{x^2} \int_0^1 du \left( f(Y_{j-1}^N(1-x) + x1\{u \leq Y_{j-1}^N\}) - f(Y_{j-1}^N) \right)$$

$$\leq C_\epsilon \|f'\|_\infty + \|f\|_\infty.$$ 

We leave to the reader the proof of the following bound:

$$\left| \int_0^1 \frac{dx}{x^2} \int_0^1 du \left( f(Y_{j-1}^N(1-x) + x1\{u \leq Y_{j-1}^N\}) - f(Y_{j-1}^N) \right) \right| \leq \epsilon \|f''\|_\infty$$

We have shown that

$$\left| q_j \mathbb{E} \left( f(Y_{j-1}^N) - f(Y_{j-1}^N) \bigg| \mathcal{F}_{\tau_j}^N \right) + \alpha Y_{j-1}^N(1 - Y_{j-1}^N)f'(Y_{j-1}^N)$$

$$\left. - \int_0^1 \frac{dx}{x^2} \int_0^1 du \left( f(Y_{j-1}^N(1-x) + x1\{u \leq Y_{j-1}^N\}) - f(Y_{j-1}^N) \right) \right|$$

$$\leq C_\epsilon (\|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty).$$

On the event $\{z > \tau_{j+1}\}$, we can replace $\Delta_j$ by $1/q_j$, modifying slightly the constant $C$, which nonetheless would not depend on $\epsilon, \delta$ and $T$. 

\[ \blacksquare \]

**Lemma 10.** Any weak limit $(\mathcal{Y}_{t \in [2, T-1]}$ of $(\mathcal{Y}_{t \in [2, T-1]}$ solves the following martingale problem

$$M_t = f(\mathcal{Y}_t) - f(y) - \int_2^t -\alpha \mathcal{Y}_v(1 - \mathcal{Y}_v)f'(\mathcal{Y}_v)dv$$

$$- \int_2^t dv \int_{|u|^2} dp \left( f(\mathcal{Y}_v + p\{u \leq \mathcal{Y}_v\} - \mathcal{Y}_v) - f(\mathcal{Y}_v) - p\{u \leq \mathcal{Y}_v\} - \mathcal{Y}_v) f'(\mathcal{Y}_v) \right)$$

\[ \text{Proof.} \] Suppose that $\phi : \mathbb{N} \to \mathbb{N}$ defines a subsequence such that $\mathcal{Y}_{\phi(N)} \to \mathcal{Y}$ in distribution, as $N \to \infty$. We define a sequence of random processes derived from the usual model with varying $\epsilon$ and $\delta$. More specifically, for all $\ell \geq 1$ let $\epsilon_\ell := 1/\ell^2$ and $\delta_\ell = O(\epsilon_\ell^2)$ such that (S) is satisfied (e.g. $1/2\ell^6$). Denote $\zeta_\ell$ the stopping time associated to the model with parameters
T, \epsilon_\ell, \delta_\ell and define \( N_\ell \) large enough such that \( \mathbb{P}(\zeta_\ell > a_{N_\ell} T) < \epsilon_\ell \) and Lemma 9 holds, and such that \((N_\ell)_{\ell \geq 1}\) is a subsequence of \( \phi \). Let \((Y_{j(2)}^{(\ell)})_{j(2) \leq j(T-1)}\) be the process stopped at time \( \zeta_\ell \), defined as previously, but with the varying parameters \( \epsilon_\ell, \delta_\ell, N_\ell \). Since it does not depend on \( \epsilon_\ell \) and \( \delta_\ell \), the corresponding non-stopped continuous-time process is simply \( \mathcal{Y}^{N_\ell} \).

Note that \( \zeta_\ell \leq a_{N_\ell} T \) for finitely many \( \ell \) almost surely, as a consequence of the choice of \( \epsilon_\ell \).

Therefore, almost surely, for all \( t \in [2, T - 1] \), it holds that

\[
\lim_{\ell \to \infty} M_t^{(\ell)} = \lim_{\ell \to \infty} \left( f(Y_{j(1)}^{(\ell)-1}) - f(Y_{j(2)}^{(\ell)-1}) - \sum_{j=j(2)}^{j(T-1)} \mathbb{E} \left( f(Y_j^{(\ell)}) - f(Y_{j(1)}^{(\ell)}) \middle| \mathcal{F}_t^{N_\ell} \right) \right) = M_t,
\]

thanks to Lemma 9.

We now show that \((M_t^{(\ell)})_{t \in [2, T - 1]}\) is a martingale with respect to its natural filtration, which will readily extend to its almost sure limit \((M_t)_{t \in [2, T - 1]}\). Let \( 2 \leq t < t + r \leq T - 1 \) and write

\[
\mathbb{E} \left( M_{t+r}^{(\ell)} - M_t^{(\ell)} \middle| (M_u^{(\ell)})_{u \leq t} \right) = \sum_{n=j(t)}^{\infty} \mathbb{P} \left( j(t + r) \wedge j(\zeta) = n \middle| (M_u^{(\ell)})_{u \leq t} \right) \times \sum_{j=j(t)}^{n-1} \mathbb{E} \left( f(Y_j^{(\ell)}) - f(Y_{j-1}^{(\ell)}) \right) \mathbb{E} \left( f(Y_j^{(\ell)}) - f(Y_{j-1}^{(\ell)}) \middle| \mathcal{F}_t^{N_\ell} \right) \left( (M_u^{(\ell)})_{u \leq t} \right).
\]

All the terms in the last sum are null, since the information given by \((M_u^{(\ell)})_{u \leq t}\) is contained in \( \mathcal{F}_t^{N_\ell} \) for all \( j \geq j(t) \). We thus deduce that \( M \) is a martingale, which entails the claim.

**Lemma 11.** Any solution of the martingale problem of Lemma 10 is a solution of the SDE (2).

Lemma 11 follows from Theorem 2.3 in [1], which addresses the question of when does a solution of a martingale problem is a solution of an associated SDE, for general Markov processes. For a more specific treatment of this question in our setting, the reader may read Section 3.3 of [1], that sketches an adaptation of an elegant duality argument from [1] (see proof of Lemma 1 therein).
References

[1] B. Bah and E. Pardoux. The $\lambda$-lookdown model with selection. *Stochastic Processes and their Applications*, 2015.

[2] Nathanaël Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009.

[3] J. Bertoin and J-F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields*, 2003.

[4] J. Bertoin and J-F. Le Gall. Stochastic flows associated to coalescent processes ii: Stochastic differential equation. *Ann. I. H. Poincaré*, 2005.

[5] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[6] E. Bolthausen and A.-S. Sznitman. On Ruelle’s probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.

[7] É. Brunet, B. Derrida, A. H. Mueller, and S. Munier. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E (3)*, 76(4):041104, 2007.

[8] Donald A. Dawson and Zenghu Li. Stochastic equations, flows and measure-valued processes. *Ann. Probab.*, 40(2):813–857, 2012.

[9] Michael M. Desai, Aleksandra M. Walczak, and Daniel S. Fisher. Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics*, 193(2):565–585, 2013.

[10] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[11] Thomas G. Kurtz. Equivalence of stochastic equations and martingale problems. In *Stochastic analysis 2010*, pages 113–130. Springer, Heidelberg, 2011.

[12] M. Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. in Appl. Probab.*, 32(4):983–993, 2000.

[13] M. Möhle. The coalescent in population models with time-inhomogeneous environment. *Stochastic Process. Appl.*, 97(2):199–227, 2002.

[14] Richard A. Neher and Oskar Hallatschek. Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences*, 110(2):437–442, 2013.
[15] Jason Schweinsberg. Rigorous results for a population model with selection I: evolution of the fitness distribution. *Electron. J. Probab.*, 22:Paper No. 37, 94, 2017.

[16] Jason Schweinsberg. Rigorous results for a population model with selection II: genealogy of the population. *Electron. J. Probab.*, 22:Paper No. 38, 54, 2017.