WRIGHT-FISHER PROCESSES WITH SELECTION AND MUTATION IN A RANDOM ENVIRONMENT

F. CORDERO\textsuperscript{1} AND G. VÉCHAMBRE\textsuperscript{2}

Abstract. Consider a bi-allelic population subject to neutral reproduction, genic selection and mutation, which is susceptible to exceptional changes in the environment. Neutral reproductions are modeled as in the classical Wright–Fisher diffusion model, mutation is parent independent and genic selection is reflected by an additional rate at which fit individuals reproduce. Moreover, changes in the environment accentuate the selective advantage of fit individuals. The evolution of the type composition is then described by a Wright–Fisher-type SDE with a jump term, modeling the effect of the environment and involving the stochastic derivative of a subordinator. Our interest in this paper is twofold: on the one side we aim to understand the influence of the environment in the type composition in the population, and on the other hand we aim to reveal the ancestral picture behind this model. The latter is described by means of an extension of the ancestral selection graph (ASG). The relation between forward and backward objects is given via duality. More precisely, we establish annealed and quenched moment dualities between the type-frequency process and the block-counting process of a variant of the ASG. As an application, we obtain a characterization of the moments of the asymptotic type distribution in the annealed and quenched form. In a similar way, we also obtain annealed and quenched results for the ancestral type distribution. In the absence of mutations, one of the types fixates and our results yield expressions for the fixation probabilities.

MSC 2010. Primary: 82C22, 92D15 Secondary: 60J25, 60J27

Keywords. Wright–Fisher diffusion, Moran model, random environment, ancestral selection graph, ancestral distribution, duality

Contents

1. Introduction 1
2. Description of the model and main results 4
3. Moran models: continuity with respect to the environment 18
4. Wright-Fisher process in random environment: existence, uniqueness and convergence 23
5. Type distribution and ancestral type distribution: the annealed case 26
6. Type distribution and ancestral type distribution: the quenched case 32
7. An application: exceptional environment in the recent past 41
Appendix A. Some definitions and remarks on the Skorokhod topology 42
Appendix B. Bounded Lipschitz metric and weak convergence 43
References 44

1. Introduction

The Wright–Fisher diffusion model with two-way mutation and fecundity selection is a classical model in population genetics. It describes the evolution forward in time of the type composition of an infinite haploid population with two types, type 0 and type 1, subject to mutations and selection. Individuals of type 0 reproduce at rate $1 + \sigma$, $\sigma \geq 0$, and individuals of type 1 reproduce at rate 1. Accordingly, we...
refer to type 0 as the fit or beneficial type and to type 1 as the unfit or deleterious type. In addition, individuals mutate at rate \( \theta \), receiving type 0 with probability \( \nu_0 \in [0, 1] \) and type 1 with probability \( \nu_1 \in [0, 1] \), with \( \nu_0 + \nu_1 = 1 \). In this model, the proportion of fit individuals evolves forward in time according to the following SDE

\[
dX(t) = [\theta \nu_0 (1 - X(t)) - \theta \nu_1 X(t) + \sigma X(t)(1 - X(t))] dt + \sqrt{2X(t)(1 - X(t))} dB(t), \quad t \geq 0,
\]

where \( (B(t))_{t \geq 0} \) is a standard Brownian motion.

The ancestral counterpart associated to this model is given by the ancestral selection graph (ASG), which traces back potential ancestors of an untyped sample of the population at present. It was introduced in the seminal work of Krone and Neuhauser \([27, 31]\), and later extended to models describing more general neutral reproduction mechanisms \([2, 15]\), or including general forms of frequency dependent selection (see \([2, 13, 12, 30]\)).

In the absence of mutations, it is known that the proportion of unfit individuals satisfy a moment duality with the block-counting process of the ASG. More precisely, if \( R(t) \) denotes the number of lineages present in the ASG at time \( t \), then

\[
\mathbb{E}_x [(1 - X(t))^n] = \mathbb{E}_n [(1 - x)^{R(t)}], \quad n \in \mathbb{N}, x \in [0, 1].
\]

As a consequence, the absorption probability of \( X \) at 0 can be expressed in terms of the probability generating function of the stationary distribution of the process \( R \). In the presence of mutations, two variants of the ASG permit to dynamically resolve mutation events and encode relevant information of the model: the killed ASG and the pruned lookdown ASG. Both processes coincide with the ASG in the absence of mutations. The killed ASG is reminiscent to the coalescent with killing \([14, \text{Chap. 1.3.1}]\) and it was introduced in \([1]\) for the Wright-Fisher diffusion with selection and mutation. The killed ASG helps to determine weather or not all the individuals in the initial sample are unfit. Moreover, its block-counting process extends the moment duality \([12]\) to this context. The latter allows to characterize the moments of the stationary distribution of \( X \) via a linear system of equations. The pruned lookdown-ASG in turn was introduced in \([28]\) for the Wright–Fisher diffusion with selection and mutation (see also \([2, 12]\) for extensions), and it helps to determine the common ancestor type-distribution.

An important assumption in the previous model is that the fitness of an individual is not affected by environmental factors. However, in many biological situations the strength of selection fluctuates in time according to changes in environmental and genetic background. The influence of random fluctuations in selection intensities on the growth of populations has been the object of extensive research in the past (see e.g. \([3, 10, 13, 24, 25, 26]\)) and it is currently experiencing renewed interest (see e.g. \([6, 8, 11, 27]\)). In this paper, we introduce a variant of the Wright–Fisher diffusion model with mutation and selection that considers the special scenario where the selective advantage of fit individuals is accentuated by exceptional environmental conditions (e.g. extreme temperatures, precipitation, humidity variation, abundance of resources, etc.). In our model the environment is described via an at most countable random collection \((t_i, p_i)_{i \in I}\) of points in \([0, \infty) \times (0, 1)\). Each \( t_i \) represents a time at which an exceptional external event occur, favoring the reproduction of the fit individuals. The value \( p_i \) models the strength of the exceptional event occurring at time \( t_i \) and we refer to it as the peak of the environment at time \( t_i \). The effect of the peak \( p_i \) is that, at time \( t_i \), each fit individual gives birth to a new fit individual with probability \( p_i \), which in turn replaces an unfit (resp. fit) individual with probability \( 1 - X(t_i -) \) (resp. \( X(t_i -) \)). Moreover, we assume that \((t_i, p_i)_{i \in I}\) is a Poisson point process on \([0, \infty) \times (0, 1)\) with intensity measure \( dt \times \mu \), where \( dt \) stands for the Lebesgue measure and \( \mu \) is a measure on \((0, 1)\) satisfying \( \int x \mu(dx) < \infty \).

Instead of working directly with the collection of peaks of the environment, it is convenient to deal with their cumulative effect: \( J(t) := \sum_{i \leq t} p_i \). The integrability condition \( \int x \mu(dx) < \infty \) implies that this sum is almost surely finite. Moreover, by the Lévy-Ito decomposition, \( (J(t))_{t \geq 0} \) is a pure jump subordinator with Lévy measure \( \mu \). From construction, the probability that a fit individual reproduces due to a peak of the environment in the infinitesimal time interval \([t, t + dt]\) is \( dJ(t) := J(t + dt) - J(t) \). Each of these reproductions lead to a replacement of an unfit individual with probability \( 1 - X(t -) \). This means that the proportion of fit individuals \( X(t -) \) is increased due to environmental events by \( X(t -)(1 - X(t -))dJ(t) \) on the infinitesimal time interval \([t, t + dt]\). Thus, the evolution of the proportion
of fit individuals is given by the SDE
\[ dX(t) = \theta (\nu_0(1 - X(t)) - \nu_1 X(t)) dt + X(t-) (1 - X(t-)) dS(t) + \sqrt{2X(t)(1 - X(t))} dB(t), \quad t \geq 0, \]
where \( S(t) := \sigma t + J(t), \ t \geq 0. \) In this SDE the Brownian motion \( B \) and the pure jump subordinator \( J \) are assumed to be independent. We refer to \( X \) as the Wright–Fisher process in random environment.

The aim of this paper is to generalize the construction of the ASG and its relatives to this model and to use them in order to characterize: (a) the asymptotic behavior of the solution of (1.3), and (b) the ancestral type distribution. We will distinguish between annealed and quenched results. In the annealed case, the environment is considered as a random object and results are stated in terms of the Lévy measure \( \mu \). In the quenched case, results are stated in terms of a fixed realization of the environment.

In order to achieve these tasks, we proceed in three steps. First, we construct a finite population analog of the previously described model having an inherent graphical representation encoding simultaneously the evolution of the forward model and the corresponding ancestral structures. In a second step, we let the size of the population converge to infinity under an appropriate scaling of parameters and time. In doing this, the forward particle picture is lost, but the frequency of type-0 individuals converges to the solution of the SDE (1.3). Moreover, the ancestral structures have natural extensions in the large population limit, giving rise to the aimed generalizations of the ASG and its relatives. In the annealed case, we establish a moment duality between the solution of (1.3) and the block-counting process of the killed ASG. As a corollary we obtain a characterization of the asymptotic type frequencies. Similarly, we characterize the common ancestor type-distribution in terms of the block-counting process of the pruned lookdown-ASG. Analogous results are obtained in the quenched case under the extra assumption that the environment is driven by a compound Poisson process.

At this point, we would like to draw the attention towards the recent work of González Casanova et al. [20], where the authors obtain a generalization of the SDE (1.3) with \( \theta = 0 \) as the large population limit of a sequence of (discrete-time) Wright–Fisher models. In their work duality also plays a crucial role. Quenched and annealed dualities are obtained at the level of finite populations. Moreover, an annealed moment duality for the jump-diffusion limit is established, which permits to elucidate a simple criteria for the accessibility of the boundaries. Unfortunately, the latter does not apply to our model. Nevertheless, we show that, for our class of models with \( \theta = 0 \), both boundaries are always accessible and we provide a characterization for the fixation probabilities.

**Outline.** The article is organized as follows. Section 2 provides an outline of the paper and contains all our main results. The proofs and more in-depth analyses are shifted to the subsequent sections. Section 3 deals with continuity properties of the type-frequency process of a Moran model with respect to the environment. In Section 4 we prove the existence and pathwise uniqueness of strong solutions of (1.3). Moreover, we show the quenched and annealed convergence of the type-frequency process of the Moran model towards the solution of (1.3). Section 5 is devoted to the proofs of the annealed results related to: (i) the moment duality between the process \( X \) and the block-counting process of the killed ASG, (ii) the characterizations of the ancestral type distribution. Section 6 deals with the quenched versions of the previous results under the assumption that the environment is driven by a compound Poisson process. We end the paper in Section 7 with an application to a model where we dispose of quenched information about the environment close to the present, but only annealed information about the environment in the far past.

**Notation.** We end the introduction with notations that will be used throughout the paper. The positive integers are denoted by \( \mathbb{N} \) and we set \( \mathbb{N}_0 := \mathbb{N} \cup \{0\} \). For \( m \in \mathbb{N} \),
\[ [m] := \{1, \ldots, m\}, \quad [m]_0 := [m] \cup \{0\}, \quad \text{and} \quad [m] := [m] \setminus \{1\}. \]
For \( N \in \mathbb{N} \), we set \( E_N := \{k/N : k \in [N]_0\} \).

For \( T > 0 \), we denote by \( \mathcal{D}_T \) (resp. \( \mathcal{D} \)) the space of càdlàg functions from \([0, T]\) (resp. \([0, \infty)\)) to \( \mathbb{R} \). For any Borel set \( S \subseteq \mathbb{R} \), denote by \( \mathcal{M}_f(S) \) (resp. \( \mathcal{M}_1(S) \)) the set of finite (resp. probability) measures on \( S \). We use \( \Rightarrow \) to denote convergence in distribution of random variables and \( \overset{(d)}{\Rightarrow} \) for convergence in distribution of càdlàg process, where the space of càdlàg functions is endowed with the Skorokhod topology.
Let \( n, m, k \in \mathbb{N}_0 \) with \( k, m \in [n]_0 \). For a random variable \( K \), we write \( K \sim \text{Hyp}(n, m, k) \) if \( K \) has a hypergeometric distribution with parameters \( n, m, \) and \( k \), i.e.

\[
P(K = i) = \frac{\binom{n-m}{i} \binom{m}{k}}{\binom{n}{k}}, \quad i \in [k \wedge m]_0.
\]

Furthermore, let \( x \in [0, 1] \) and \( n \in \mathbb{N} \). For a random variable \( B \), we write \( B \sim \text{Bin}(n, x) \) if \( B \) has a binomial distribution with parameter \( n \) and \( x \), i.e.

\[
P(B = i) = \binom{n}{i} x^i (1 - x)^{n-i}, \quad i \in [n]_0.
\]

2. Description of the model and main results

In this section we provide a detailed outline of the paper and state the main results.

2.1. Moran models in deterministic pure jump environments.

2.1.1. The model. We consider a haploid population of size \( N \) with two types, type 0 and type 1, subject to mutation and random reproduction, that evolves in a deterministic environment.

The environment is modeled by an at most countable collection \( e := (t_k, p_k)_{k \in I} \) of points in \((\mathbb{R}, \mathbb{N}) \times (0, 1)\) satisfying that \( \sum_{k,t_k \in [s,t]} p_k < \infty \), for any \( s \leq t \). We refer to \( p_k \) as the peak of the environment at time \( t_k \).

The individuals in the population undergo the following dynamic. Each individual mutates at rate \( \sigma_N \) independently of each other; acquiring type 0 (resp. 1) with probability \( \nu_0 \) (resp. \( \nu_1 \)), where \( \nu_0, \nu_1 \in [0, 1] \) and \( \nu_0 + \nu_1 = 1 \). Reproduction occurs independently of mutation, individuals of type 1 reproduce at rate 1, whereas individuals of type 0 reproduce at rate \( 1 + \sigma_N \). \( \sigma_N \geq 0 \). In addition, at time \( t_k, k \in I \), each type 0 individual reproduces with probability \( p_k \), independently from the others. At any reproduction time: (a) each individual produces at most one offspring, which inherits the parent’s type, and (b) if \( n \) individuals are born, \( n \) individuals are randomly sampled without replacement from the extant population to die, hence keeping the size of the population constant.

2.1.2. Graphical representation. In the absence of environmental factors (i.e. \( e = \emptyset \)), the evolution of the population is commonly described by means of its graphical representation as an interactive particle system. The latter allows to decouple the randomness of the model coming from the initial type configuration and the one coming from mutations and reproductions. Such a construction can be extended to include the effect of the environment as follows. Non-environmental events are as usual encoded via a family of independent Poisson processes

\[
\Lambda := \{(\lambda^0_i, \lambda^1_i) : \{\lambda^0_{i,j}, \lambda^1_{i,j}\}_{j \in [N]} \}_i \in [N],
\]

where: (a) for each \( i, j \in [N] \) with \( i \neq j \), \((\lambda^0_{i,j}(t))_{t \in \mathbb{R}}\) and \((\lambda^1_{i,j}(t))_{t \in \mathbb{R}}\) are Poisson processes with rates \( \sigma_N/N \) and \( 1/N \), respectively, and (b) for each \( i \in [N] \), \((\lambda^0_i(t))_{t \in \mathbb{R}}\) and \((\lambda^1_i(t))_{t \in \mathbb{R}}\) are Poisson processes with rates \( \theta_N \nu_0 \) and \( \theta_N \nu_1 \), respectively. We call \( \Lambda \) the basic background. The environment introduces a new independent source of randomness into the model, that we describe via the collection

\[
\Sigma := \{(U_i(t))_{t \in [N], t \in \mathbb{R}}, (\tau_A(t))_{A \in [N], t \in \mathbb{R}}\}
\]

where: (c) \((U_i(t))_{t \in [N], t \in \mathbb{R}}\) is a \([N] \times \mathbb{R}\)–indexed family of i.i.d. random variables with \( U_i(t) \) being uniformly distributed on \([0, 1]\), and (d) \((\tau_A(t))_{A \in [N], t \in \mathbb{R}}\) is a family of independent random variables with \( \tau_A(t) \) being uniformly distributed on the set of injections from \( A \) to \([N] \). We call \( \Sigma \) the environmental background. We assume that basic and environmental backgrounds are independent and we call \((\Lambda, \Sigma)\) the background.

In the graphical representation of the Moran model individuals are represented by horizontal lines at the integer levels in \([N]\). Time runs from left to right. Each reproduction event is represented by an arrow with the parent at its tail and the offspring at its head. We distinguish between neutral reproductions, depicted as filled head arrows, and selective reproductions, depicted as open head arrows. Mutation events are depicted by crosses and circles on the lines. A circle (cross) indicates a mutation to type 0 (type 1). See Fig. 1 for a picture. The appearance of all these random elements can be read off from the
background as follows. At the arrival times of $\lambda^i_{k,j}$ (resp. $\lambda^i_{k,j}$), we draw open (resp. filled) head arrows from level $i$ to level $j$. At the arrival times of $\lambda^0_i$ (resp. $\lambda^i_1$), we draw an open circle (resp. a cross) at level $i$. In order to draw the environmental elements, we define, for each $k \in I$,

$$I_e(k) := \{ i \in [N] : U(s_i(t_k) \leq p_k \} \quad \text{and} \quad n_e(k) := |I_e(k)|$$

and at time $t_k$, we draw for any $i \in I_e(k)$ an open arrow head from level $i$ to level $\tau_{I_e(k)}(i)$.

**Remark 2.1.** Note that, for any $s < t$, the number of basic graphical elements present in $[s,t]$ is almost surely finite. Moreover, since

$$E \left[ \sum_{k: t_k \in [s,t]} n_e(k) \right] = N \sum_{k: t_k \in [s,t]} p_k < \infty,$$

and hence $\sum_{k: t_k \in [s,t]} n_e(k) < \infty$ almost surely. This means that the number of arrows in $[s,t]$ due to peaks of the environment is almost surely finite.

Now, given a realization of the particle system and an assignment of types to the lines at time $s$, we propagate types along the lines forward in time respecting mutations and reproduction events (see Fig 1). By this we mean that, as long as we move from left to right in the graphical picture, the type of a given line remains unchanged until the line encounters a cross, a circle or an arrowhead. If the line encounters a cross (resp. a circle) at time $t$, it gets type 1 (resp. type 0) from time $t$ until the next cross, circle or arrowhead. If at time $t$ the line encounters a neutral arrowhead, the line gets at time $t$ the type of the line at the tail of the arrow. The same happens if the line encounters a selective arrowhead and the line at its tail has type 0, otherwise the selective arrow is ignored.

**Remark 2.2.** One of the advantages of using such a graphical representation is that using the same background one can couple Moran models with the same parameters of selection and mutation but evolving in a different environment, or or two Moran models with the same parameters and environment, but with a different initial composition of types.

### 2.1.3. Type frequency process

In what follows, we assume that we know the type composition at time $s = 0$ and that that there is no peak at that time. As mentioned in the introduction it is convenient to work with the cumulative effect of the peaks rather than with the peaks themselves. Thus, we define the function $\omega : [0,\infty) \to \mathbb{R}$ via

$$\omega(t) := \sum_{e: t_e \in [0, t]} p_e, \quad t \geq 0.$$

The function $\omega$ is càdlàg, non-decreasing and satisfies

1. for all $t \in [0,\infty)$, $\Delta \omega(t) := \omega(t) - \omega(t- \in [0,1)$ and $\sum_{u \in [0,t]} \Delta \omega(u) < \infty$;
2. $\omega(0) = 0$ and $\omega$ is pure jump, i.e. for all $t \in [0,\infty)$, $\omega(t) = \sum_{u \in [0,t]} \Delta \omega(u)$.

We denote by $\mathbb{D}^*$ the set of all càdlàg, non-decreasing functions satisfying (i) and (ii). Note that the environment in $[0,\infty)$ is given by the collection of points $\{(t, \Delta \omega(t)) : \Delta \omega(t) > 0\}$. Hence, the function $\omega$ completely determines the environment in $[0,\infty)$. For this reason, we often abuse of the notation and
refer to $\omega$ as the environment. In addition, an environment $\omega \in \mathbb{D}^*$ is said to be simple if $\omega$ has only a finite number of jumps in any compact time interval.

For any $t \geq 0$, we denote by $X_N(\omega, t)$ the proportion of fit individuals at time $t$ in the population. We denote by $\mathbb{P}_N^\omega$ the law of the process $X_N(\omega, \cdot)$, and we refer to it as the quenched probability measure.

For the null environment, i.e., $\omega \equiv 0$, $X_N(0, \cdot)$ is the continuous-time Markov chain on $E_N$ with infinitesimal generator

$$A_N^0f(x) = (N(1 + \sigma_N)x(1 - x) + N\theta_N\nu_0(1 - x)) \left( f \left( x + \frac{1}{N} \right) - f(x) \right) + (Nx(1 - x) + N\theta_N\nu_1x) \left( f \left( x - \frac{1}{N} \right) - f(x) \right), \quad x \in E_N.$$ 

For a simple environment $\omega$ with jumping times $s_1 < \cdots < s_k$ in $[0, T]$, the evolution of $X_N(\omega, \cdot)$ is as follows. If $X_N(\omega, 0) = x_0 \in E_N$, then $X_N(\omega, \cdot)$ evolves in $[0, s_1)$ as the continuous-time Markov chain with infinitesimal generator $A_N^0$ starting at $x_0$. Similarly, in the intervals $[s_i, s_{i+1})$, $X_N(\omega, \cdot)$ evolves as the continuous-time Markov chain with infinitesimal generator $A_N^0$ starting at $X_N(\omega, s_i)$. Moreover, if $X_N(\omega, s_i) = x$, then $X_N(\omega, s_i) = x + \frac{H_i(x, \mathcal{B}_i(x))}{N}$, where the random variables $H_i(x, n) \sim \text{Hyp}(N, N(1 - x), n)$, $n \in [N]_0$, and $\mathcal{B}_i(x) \sim \text{Bin}(Nx, \Delta\omega(s_i))$ are independent.

We describe now the dynamic of the type frequency process in the case of a general environment $\omega$. Let $(S_i)_{i \in \mathbb{N}}$ be the increasing sequence of times at which environmental reproductions take place, and set $S_0 := 0$. From construction this sequence is Markovian and its transition probabilities are given by

$$\mathbb{P}_N^\omega(S_{i+1} > t \mid S_i = s) = \prod_{u \in (s, t]} (1 - \Delta\omega(u))^N, \quad i \in \mathbb{N}_0, 0 \leq s \leq t.$$ 

If $X_N(\omega, 0) = x_0 \in E_N$, then $X_N(\omega, \cdot)$ evolves in $[0, S_1)$ as the continuous-time Markov chain with infinitesimal generator $A_N^0$ starting at $x_0$. In the intervals $[S_i, S_{i+1})$, $X_N(\omega, \cdot)$ evolves as the continuous-time Markov chain with infinitesimal generator $A_N^0$ starting at $X_N(\omega, S_i)$. Moreover, if $X_N(\omega, S_i) = x$, then $X_N(\omega, S_i) = x + \frac{H_i(x, \mathcal{B}_i(x))}{N}$, where the random variables $H_i(x, n) \sim \text{Hyp}(N, N(1 - x), n)$, $n \in [N]_0$, and $\mathcal{B}_i(x)$ are independent, and $\mathcal{B}_i(x)$ has the distribution of a binomial random variable with parameters $Nx$ and $\Delta\omega(s_i)$ conditioned to be positive.

We end this section with a result that provides the continuity of the type-frequency process with respect to the environment. For this we restrict ourselves to realizations of the model to the time interval $[0, T]$. Note that the restriction of the environment to $[0, T]$ can be identified with an element of

$$\mathbb{D}_T^* := \{ \omega \in \mathbb{D}_T : \omega(0) = 0, \Delta\omega(t) \in [0, 1) \text{ for all } t \in [0, T], \omega \text{ is non-decreasing and pure-jump} \}.$$ 

Moreover, we equip $\mathbb{D}_T^*$ with the metric $d_T^*$ defined in Appendix A.

**Theorem 2.1** (Continuity). Let $\omega \in \mathbb{D}_T^*$ and let $\{\omega_k\}_{k \in \mathbb{N}} \subset \mathbb{D}_T^*$ be such that $d_T^*(\omega_k, \omega) \to 0$ as $k \to \infty$. If $X_N(\omega_k, 0) = X_N(\omega, 0)$ for all $k \in \mathbb{N}$, then

$$(X_N(\omega_k, t))_{t \in [0, T]} \xrightarrow{d_T^*} (X_N(\omega, t))_{t \in [0, T]},$$

2.2. Moran models in a environment driven by a subordinator. In contrast to Section 2.1, we consider here a random environment given by a Poisson point process $(t_i, p_i)_{i \in \mathbb{E}}$ on $[0, \infty) \times (0, 1)$ with intensity measure $dt \times \mu$, where $dt$ stands for the Lebesgue measure and $\mu$ is a measure on $(0, 1)$ satisfying $\int x \mu(dx) < \infty$. The integrability condition on $\mu$ implies that, for every $t \geq 0$, $J(t) := \sum_{i \leq t} p_i < \infty$ almost surely. Moreover, $J \in \mathbb{D}_*^\mu$ almost surely. Note that by the Lévy-Itô decomposition, $(J(t))_{t \geq 0}$ is a pure jump subordinator with Lévy measure $\mu$. If the measure $\mu$ is finite, then $J$ is a compound Poisson process so the environment $J$ is almost surely simple.

**Theorem 2.1** together with the graphical representation of the Moran model given in Section 2.1.2 provide a natural way to construct such a model. Indeed, on the basis of a common background $(\Lambda, \Sigma)$, we can construct simultaneously Moran models for any $\omega \in \mathbb{D}_*^\mu$. Next, we consider a pure jump subordinator $(J(t))_{t \geq 0}$ independent of the background, with Lévy measure $\mu$ on $(0, 1)$ satisfying $\int x \mu(dx) < \infty$. 


Theorem 2.1 assures then that the process \((X_N(J, t))_{t \geq 0}\) is well defined. Its law \(P_N\) is called the annealed probability measure. The formal relation between the annealed and quenched measures is given by
\[
P_N(\cdot) = \int P_N^\omega(\cdot) P(\omega),
\]
where \(P\) denotes the law of \(J\).

By construction the process \((X_N(J, t))_{t \geq 0}\) is a continuous time Markov chain on \(E_N\) and its infinitesimal generator \(A_N\) is given by
\[
A_N f(x) := A_N^0 f(x) + \int_{(0,1)} \left( E \left[ f \left( x + \frac{\mathcal{H}(N, N(1-x), \beta(Nx, u))}{N} \right) \right] - f(x) \right) \mu_N(du), \quad x \in E_N,
\]
where \(\beta(Nx, u) \sim \text{Bin}(Nx, u)\), and for any \(i \in [Nx]_0\), \(\mathcal{H}(N, N(1-x), i) \sim \text{Hyp}(N, N(1-x), i)\) are independent.

The dynamic of the graphical representation is as follows: For each \(i, j \in [N]\) with \(i \neq j\), open (resp. filled) head arrows from level \(i\) to level \(j\) appear at rate \(\sigma_{N, i} / N\) (resp. \(1/N\)). For each \(i \in [N]\), open circles (resp. crosses) appear at level \(i\) at rate \(\theta_{N, i} \sim \text{Bin}(N, u_i)\) (resp. \(\theta_{N, i}\)). For each \(k \in [N]\), every group of \(k\) lines is subject to a simultaneous reproduction at rate
\[
\sigma_{N, k} := \int_{(0,1)} y^k (1-y)^{N-k} \mu(dy).
\]
In such a simultaneous reproduction, the set of the \(k\) descents is chosen uniformly at random among the \(N\) individuals, and \(k\) open head arrows are drawn uniformly at random from the \(k\) parents to the \(k\) descents.

2.3. The Wright–Fisher process in random environment. In this section we are interested in the Wright–Fisher process in random environment described in the introduction. We start stating the well-posedness of the SDE \((\ref{eq:Wright-Fisher})\).

**Theorem 2.2** (Existence and uniqueness). Let \(\sigma, \theta \geq 0, \nu_0, \nu_1 \in [0, 1]\) with \(\nu_0 + \nu_1 = 1\). Let \(J\) be a pure jump subordinator with \(\mu\) supported in \((0, 1)\) and let \(B\) be a standard Brownian motion independent of \(J\). Then, for any \(x_0 \in [0, 1]\), there is a pathwise unique strong solution \((X(t))_{t \geq 0}\) to the SDE \((\ref{eq:Wright-Fisher})\) such that \(X(0) = x_0\). Moreover, \(X(t) \in [0, 1]\) for all \(t \geq 0\).

**Remark 2.3.** The Wright–Fisher process defined via the SDE \((\ref{eq:Wright-Fisher})\) with \(\theta = 0\) corresponds to \(\cite{20}\), Eq. (10) with \(K, y \in (0, 1)\), being a random variable that takes the value 1 with probability \(1-y\) and the value 2 with probability \(y\).

Let us now consider \(J = (J(s))_{s \geq 0}\) and \(B = (B(s))_{s \geq 0}\) as in the previous theorem. Therefore, for any \(t > 0\), the solution of \((\ref{eq:Wright-Fisher})\) in the time interval \([0, t]\) is a measurable function of \((B(s), J(s))_{s \in [0,t]}\), which we denote by \(F(B, J)\). We denote by \(P^{\omega}\) a regular version of the conditional law of \(F(B, J)\) given \(J\), which is called the quenched probability measure. This allows us to write \(P^{\omega}\) for a.e. realization \(\omega\) of \(J\).

If \(P\) denotes the law of \(J\), we called the **annealed measure**
\[
P(\cdot) = \int P^{\omega}(\cdot) P(\omega).
\]
We write \((X(t))_{t \geq 0}\) for the solution of \((\ref{eq:Wright-Fisher})\) under the annealed measure. Similarly, we write \((X(\omega, t))_{t \geq 0}\) for the solution of \((\ref{eq:Wright-Fisher})\) under the quenched measure \(P^{\omega}\). The next result provides the convergence, under a suitable scaling of parameters and of time, of the annealed measures of a sequence of Moran models to the annealed measure of the Wright–Fisher process in random environment.

**Theorem 2.3** (Annealed convergence). Let \(J\) be a pure jump subordinator with \(\mu\) supported in \((0, 1)\), and set \(J_N(t) := J(t/N), t \geq 0\). Assume in addition that
1. \(\sigma_{N} \rightarrow \sigma\) and \(\theta_{N} \rightarrow \theta\) for some \(\sigma \geq 0, \theta \geq 0\) as \(N \rightarrow \infty\).
2. \(X_N(J_N, 0) \rightarrow x_0\) as \(N \rightarrow \infty\).
Then, we have

\[(X_N(J_N, Nt))_{t \geq 0} \xrightarrow{\mathbb{D}^\omega} (X(t))_{t \geq 0},\]

where \(X\) is the unique pathwise solution of (1.3) with \(X(0) = x_0\).

For \(\omega\) simple, the quenched law \(\mathbb{P}^\omega\) can be alternatively defined as follows. Denote by \(s_1 < \cdots < s_k\) the jumps of \(\omega\) in \([0, T]\). If \(X(\omega, 0) = x_0\), \(X(\omega, \cdot)\) evolves in \([0, s_1]\) as the solution of (1.1) starting at \(x_0\). In the intervals \([s_i, s_{i+1})\), \(X_N(\omega, \cdot)\) evolves as the solution of (1.1) starting at \(X(\omega, s_i)\). Moreover, if \(X(\omega, s_i-) = x\), then \(X(\omega, s_i) = x + x(1-x)\Delta \omega(s_i)\). The next result extends Theorem 2.3 to the quenched setting for simple environments.

**Theorem 2.4 (Quenched convergence).** Let \(\omega \in \mathbb{D}^\tau\) be a simple environment and set \(\omega_N(t) = \omega(t/N)\), \(t \geq 0\). Assume in addition that

1. \(N\sigma_N \to \sigma\) and \(N\theta_N \to \theta\) for some \(\sigma \geq 0, \theta \geq 0\) as \(N \to \infty\).
2. \(X_N(\omega_N, 0) \to X(\omega, 0)\) as \(N \to \infty\).

Then, we have

\[(X_N(\omega_N, Nt))_{t \geq 0} \xrightarrow{\mathbb{D}^\omega} (X(\omega, t))_{t \geq 0}\]

**Remark 2.4.** Recall that if \(J\) is a compound Poisson process, then almost every environment is simple. Hence, in this case, Theorem 2.4 tells us that the quenched convergence holds for \(P\)-almost every environment. We conjecture that this statement holds true in the general case where \(J\) is a subordinator with Lévy measure supported in \((0, 1)\). The main difficulty in proving such a result comes from the fact that the diffusion term in (1.3) is not Lipschitz.

### 2.4. The ancestral selection graph.

The aim of this section is to generalize the construction of the classical ancestral selection graph (ASG) of Krone and Neuhauser [27, 31] to include the effect of the random environment. In order to motivate our definition, we briefly explain in Section 2.4.1 how to read off the ASG in a Moran model in deterministic pure jump environment on the basis of its graphical representation. Inspired in this construction, we define in Section 2.4.2 the ASG for the Wright–Fisher process in random environment under the annealed measure. For simple environments, we also define the quenched version of the ASG. The latter can be extended to the general case, but in order to keep things simple and since anyways our results (in the quenched case) are stated only in the case of simple environments, we refrain to do this here.

#### 2.4.1. Reading off the ancestors in the Moran model.

The ancestral selection graph (ASG) was introduced by Krone and Neuhauser in [27] (see also [31]) to study the genealogical relations of an untyped sample taken from the population at present, in the diffusion limit of the Moran model with mutation and selection. In this section we adapt this construction to the Moran model in deterministic pure jump environment described in Section 2.1.1.

Let us fix a simple environment and consider a realization of the interactive particle system associated to the Moran model described in Section 2.1.1 in the time interval \([0, T]\). We aim to trace backward in time (from right to left in Figure 2) the lines of the potential ancestors of \(n\) sampled individuals at time \(T\). In particular, the timelines of the ASG and of the type-frequency process run in the opposite direction, the backward time \(\beta \in [0, T]\) corresponding to the forward time \(t = T - \beta\). So, we start at time \(\beta = 0\) with the \(n\) lines corresponding to the \(n\) sampled individuals and we follow their lines of descent in the backward direction of time as follows.

When a neutral arrow joins two individuals in the current set of potential ancestors, the two lines coalesce into a single one, the one at the tail of the arrow. When a neutral arrow hits from outside a potential ancestor, the hit line is replaced by the line at the tail of the arrow. When a selective arrow hits the current set of potential ancestors, the individual that is hit has two possible parents, the incoming branch at the tail and the continuing branch at the tip. The true parent depends on the type of the incoming branch, but for the moment we work without types. These unresolved reproduction events can be of two types: a branching event if the selective arrow emanates from an individual outside the current set of potential ancestors, and a collision event if the selective arrow links two current potential ancestors. Note that at the jumping times of the environment, multiple
lines in the ASG can be hit by selective arrows, and therefore, multiple branching and collision events may occur simultaneously. Mutations remain superposed on the lines of the ASG. From construction the number of potential ancestors decreases by one in a coalescence event, it is increased by one at any (local) branching event, and remains unchanged in collision and replacement events (see Fig. 2). The ASG in $[0, T]$ contains all the lines that are potentially ancestral (ignoring mutation events) to the sampled lines at time $t = T$. Given an assignment of types to the lines present in the ASG at time $t = 0$, we can extract the true genealogy and determine the types of the sampled individuals at time $T$. For this, we propagate types forward in time along the lines of the ASG taking into account mutations and reproductions, with the rule that if a line is hit by a selective arrow, the incoming line is the ancestor if and only if it is of type 0, see Figure 3. This rule is called the pecking order. Proceeding in this way, the types in $[0, T]$ are determined along with the true genealogy.

Remark 2.5. Using the dynamic of the graphical representation of the Moran model, that is described in Section 2.1 for the quenched case and in Section 2.2 for the annealed case, it is not difficult to deduce from above the dynamic of the ASG of the Moran model in both the quenched and annealed cases.

2.4.2. The ASG for the Wright–Fisher process in random/fixed environment. The aim of this section is to extend the definition of the ancestral selection graph to the large population limit described in the previous section. In contrast to the construction given in Section 2.1.1 for the Moran model, we do not dispose of a graphical representation for the Wright–Fisher process in random environment. One option would be to provide such a graphical construction in the spirit of the lookdown construction of [4]. However, we opt here for the following alternative strategy. We first consider the graphical representation of a Moran model with parameters $\sigma/N$, $\theta/N$, $\nu_0$, $\nu_1$ and environment $\omega_N(\cdot) = \omega(\cdot/N)$, and we speed up time by $N$. We know that the type-0 frequency process will converge as $N \to \infty$ to the quenched solution of [1]. Now, we sample $n$ individuals at time $T$ and we construct the ASG as in Section 2.4.1. A simple asymptotic analysis of the rates and probabilities for the possible events leads to the following definitions.

Definition 2.5 (The quenched ASG in a simple environment). Let $\omega : \mathbb{R} \to \mathbb{R}$ be a simple environment. The quenched ancestral selection graph with parameters $\sigma, \theta, \nu_0, \nu_1$ and environment $\omega$, associated to a sample of the population of size $n$ at time $T$ is the branching-coalescing particle system $(G^T(\omega, \beta))_{\beta \geq 0}$ starting at $\beta = 0$ with $n$ lines and with the following dynamic.
are all unfit if and only if all the lines present at time 

We can infer directly that all the sampled individuals are unfit. In the remaining case, the sampled individuals 

If in the pruned ASG there is a line ending in a beneficial mutation, then we can infer that at least one 

absence of mutations, the 

would then expect to have 

Mutations determine the types of some of the lines in the ASG even before we assign types to the lines 

environment driven by a pure jump subordinator with Lévy measure \( \mu \), associated to a sample of the population of size \( n \) at time \( T \) is the branching-coalescing particle system \( \mathcal{G}(\beta) \) starting with \( n \) lines and with the following dynamic.

- each line splits into two lines, an incoming line and a continuing line, at rate \( \sigma \).
- every given pair of lines coalesce into a single line at rate \( \beta \). 
- if at time \( \beta \), we have \( \Delta \omega(T - \beta) > 0 \), then any line splits into two lines, an incoming line and a 
  - each line is decorated by a beneficial mutation at rate \( \theta \nu_0 \).
  - each line is decorated by a deleterious mutation at rate \( \theta \nu_1 \).

**Definition 2.6** (The annealed ASG). The annealed ancestral selection graph with parameters \( \sigma, \theta, \nu_0, \nu_1 \), and environment driven by a pure jump subordinator with Lévy measure \( \mu \), associated to a sample of the population of size \( n \) at time \( T \) is the branching-coalescing particle system \( \mathcal{G}(\beta) \) starting with \( n \) lines and with the following dynamic.

- each line splits into two lines, an incoming line and a continuing line, at rate \( \sigma \).
- every given pair of lines coalesce into a single line at rate \( \beta \).
- every group of \( k \) lines is subject to a simultaneous branching at rate 
  \[
  \sigma_{m,k} := \int_{(0,1)} y^k (1-y)^{m-k} \mu(dy),
  \]
  where \( m \) denotes the total number of lines in the ASG before the *simultaneous branching event.*
  At the simultaneous branching event, each line in the group involved splits into two lines, an 
  incoming line and a continuing line.
- each line is decorated by a beneficial mutation at rate \( \theta \nu_0 \).
- each line is decorated by a deleterious mutation at rate \( \theta \nu_1 \).

**Remark 2.6.** Note that in the annealed case, the environment is homogeneous and its law in the backward and forward directions of time is the same. In particular, the distribution of the ASG does not depend on the sampling time \( T \), which explains why we omit it in the notation.

2.5. **Type frequency via the killed ASG: a moment duality.** The aim of this section is to relate the type-0 frequency process \( X \) to the ancestral selection graph. We start with an heuristic reasoning that leads to the definitions of the killed ASG in the annealed and quenched case respectively. Let us assume that the proportion of fit individuals at time 0 is equal to \( x \in [0,1] \). Conditionally on \( X(T) \), the probability of sampling independently \( n \) unfit individuals at time \( t \) equals \( (1-X(T))^n \). Now, consider the annealed ASG associated to the \( n \) sampled individuals in \([0,T]\) and assign randomly types independently to each line present in the ASG at time \( \beta = T \) according to the initial distribution \((x, 1-x)\). In the absence of mutations, the \( n \) sampled individuals are unfit if and only if all the lines in the ASG at time \( \beta = T \) are assigned the unfit type (since at any selective event a fit individual can only be replaced by another fit individual). Therefore, if \( R(T) \) denotes the number of lines present in the ASG at time \( \beta = t \), then conditionally on \( R(T) \), the probability that the \( n \) sampled individuals are unfit is \((1-x)^{R(T)}\). We would then expect to have 

\[
\mathbb{E}[(1-X(T))^n \mid X(0) = x] = \mathbb{E}[(1-x)^{R(T)} \mid R(0) = n].
\]

Mutations determine the types of some of the lines in the ASG even before we assign types to the lines at time \( T \). Hence, we can pruned away from the ASG all the sub-ASGs arising from a mutation event. If in the pruned ASG there is a line ending in a beneficial mutation, then we can infer that at least one of the sampled individuals has the fit type. If all the lines end up in a deleterious mutation, then we can infer directly that all the sampled individuals are unfit. In the remaining case, the sampled individuals are all unfit if and only if all the lines present at time \( T \) in the pruned ASG are assigned the unfit type. This motivates the following definition.

**Definition 2.7** (The annealed killed ASG). The annealed killed ASG with parameters \( \sigma, \theta, \nu_0, \nu_1 \), and environment driven by a pure jump subordinator with Lévy measure \( \mu \), associated to a sample of the population of size \( n \) is the branching-coalescing particle system \( \mathcal{G}_1(\beta) \) starting with \( n \) lines and with the following dynamic.

- each line splits into two lines, an incoming line and a continuing line, at rate \( \sigma \).
Then the diffusion Theorem 2.10 following result is a consequence of the previous duality relation.

Theorem 2.9 stated as follows.

Definition 2.8 (The quenched killed ASG for a simple environment). Let \( \omega : \mathbb{R} \to \mathbb{R} \) be a simple environment. The quenched killed ASG with parameters \( \sigma, \theta, \nu \), and environment \( \omega \), associated to a sample of the population of size \( n \) at time \( T \) is the branching-coalescing particle system \( (G^T_1(\omega, \beta))_{\beta \geq 0} \) starting at \( \beta = 0 \) with \( n \) lines and with the following dynamic.

- each line splits into two lines, an incoming line and a continuing line, at rate \( \sigma \).
- every given pair of lines coalesces into a single line at rate \( 2 \).
- if at time \( \beta \), we have \( \Delta \omega(T - \beta) > 0 \), then any particle lines splits into two lines, an incoming line and a continuing line, with probability \( \Delta \omega(T - \beta) \), independently from the other lines.
- each line is killed at rate \( \theta \nu \).
- each line sends the process to the cemetery state \( \dagger \) at rate \( \theta \nu \).

A similar reasoning leads, in the quenched case, to the following definition.

Definition 2.8 (The quenched killed ASG for a simple environment). Let \( \omega : \mathbb{R} \to \mathbb{R} \) be a simple environment. The quenched killed ASG with parameters \( \sigma, \theta, \nu \), and environment \( \omega \), associated to a sample of the population of size \( n \) at time \( T \) is the branching-coalescing particle system \( (G^T_1(\omega, \beta))_{\beta \geq 0} \) starting at \( \beta = 0 \) with \( n \) lines and with the following dynamic.

- each line splits into two lines, an incoming line and a continuing line, at rate \( \sigma \).
- every given pair of lines coalesces into a single line at rate \( 2 \).
- if at time \( \beta \), we have \( \Delta \omega(T - \beta) > 0 \), then any particle lines splits into two lines, an incoming line and a continuing line, with probability \( \Delta \omega(T - \beta) \), independently from the other lines.
- each line is killed at rate \( \theta \nu \).
- each line sends the process to the cemetery state \( \dagger \) at rate \( \theta \nu \).

The formal annealed and quenched relations between the type-frequency process and the killed ASG are established in the subsequent sections.

2.5.1. The annealed case. We start this section with the duality relation between the process \( X \) and the block-counting process of the killed ASG.

For each \( \beta \geq 0 \), we denote by \( R(\beta) \) the number of lines present in the killed ASG at time \( \beta \), with the convention that \( R(\beta) = \dagger \) if \( G_1(\beta) = \dagger \). The process \( R := (R(\beta))_{\beta \geq 0} \), called the block-counting process of the killed ASG, is a continuous-time Markov chain with values on \( \mathbb{N}_0 := \mathbb{N} \cup \{\dagger\} \) and infinitesimal generator matrix \( Q^R := (q^R(i,j))_{i,j \in \mathbb{N}_0} \) defined via

\[
q^R_{ij} := \begin{cases} 
  i(i-1) + \theta \nu_1 & \text{if } j = i - 1, \\
  i(\sigma + \sigma_1,1) & \text{if } j = i + 1, \\
  \binom{i}{k} \sigma_{i,k} & \text{if } j = i + k, \ i \geq k \geq 2, \\
  -i(i-1 + \theta + \sigma) - \int_{(0,1)} (1 - (1-y)\mu(dy)) & \text{if } j = \dagger, \\
  0 & \text{otherwise}.
\end{cases}
\] (2.3)

Note that, when \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0, 1) \), 0 and \( \dagger \) are absorbing states of \( R \). The moment duality can be stated as follows.

Theorem 2.9 (Annealed moment duality). For all \( x \in [0, 1] \) and \( n \in \mathbb{N} \)

\[
\mathbb{E}[(1 - X(T))^n | X(0) = x] = \mathbb{E}[(1 - x)^{R(T)} | R(0) = n] \tag{2.4}
\]

with the convention \((1 - x)^\dagger = 0 \) for all \( x \in [0, 1] \).

When \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0, 1) \), the absorption probabilities of \( R \) at 0 and \( \dagger \) are well-defined. In this case, for \( n \in \mathbb{N}_0 \), we set

\[
w_n := \mathbb{P}(\exists \beta \geq 0 : R(\beta) = 0 \mid R(0) = n),
\]

the probability that \( R \) is eventually absorbed at 0 conditionally on starting at \( n \). Clearly \( w_0 = 1 \). The following result is a consequence of the previous duality relation.

Theorem 2.10 (Asymptotic type-frequency: the annealed case). Assume that \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0, 1) \). Then the diffusion \( X \) has a unique stationary distribution \( \pi_X \in \mathcal{M}_1([0, 1]) \). Let \( X(\infty) \) be a random
variable distributed according to \( \pi_X \). Then \( X(t) \) converges in distribution to \( X(\infty) \), independently of the starting value of \( X \). Moreover, for all \( n \in \mathbb{N} \), we have
\[
\mathbb{E} \left[ (1 - X(\infty))^n \right] = w_n, \tag{2.5}
\]
and the absorption probabilities \( (w_n)_{n \geq 0} \) satisfy
\[
(\sigma + \theta + n - 1)w_n = \sigma w_{n+1} + (\theta \nu_1 + n - 1)w_{n-1} + \sum_{k=1}^{n} \binom{n}{k} \sigma_{n,k}(w_{n+k} - w_n), \quad n \in \mathbb{N}. \tag{2.6}
\]

### 2.5.2. The quenched case.

Let us fix \( T \in \mathbb{R} \) and a simple environment \( \omega \) on \([-\infty, T]\). We first state the moment duality between the process \( X(\omega, \cdot) \) and the quenched killed ASG \( G_T^\omega(\omega, \cdot) \). Let \( (R_T(\omega, \beta'))_{\beta \geq 0} \) be the block-counting process of the killed ASG \( G_T^\omega(\omega, \cdot) \). From construction, \( (R_T(\omega, \beta'))_{\beta \geq 0} \) is a continuous-time Markov process with values on \( \mathbb{N}_0 \). Between the jumping times of the environment \( \omega \), \( (R_T(\omega, \beta'))_{\beta \geq 0} \) has transitions with rates \( (q_T(i,j))_{i,j \in \mathbb{N}_0} \) (i.e. \( \mu = 0 \) in (2.3)). In addition, at each jumping time of the environment \( \omega \), i.e. at each time \( \beta \geq 0 \) such that \( \Delta \omega(T - \beta) > 0 \), we have
\[
\forall i \in \mathbb{N}, \forall k \in [i]_0, \quad \mathbb{P}^\omega(R_T(\omega, \beta) = i + k \mid R_T(\omega, \beta -) = i) = \binom{i}{k} (\Delta \omega(T - \beta))^k (1 - \Delta \omega(T - \beta))^{i-k}.
\]
In other words, conditionally on \( \{R_T(\omega, \beta -) = i\} \), \( R_T(\omega, \beta) \sim i + Y \) where \( Y \sim \text{Bin}(i, \Delta \omega(T - \beta)) \). As in the annealed case, \( 0 \) and \( \dagger \) are absorbing states of the killed ASG when \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0, 1) \). The moment duality can be stated as follows.

**Theorem 2.11** (Quenched moment duality). For all \( T > 0 \), \( x \in [0, 1] \), \( n \in \mathbb{N} \), we have
\[
\mathbb{E}^{\omega} \left[ (1 - X(\omega, T))^n \mid X(\omega, 0) = x \right] = \mathbb{E}^{\omega} \left[ (1 - x)^{R_T(\omega, T -)} \mid R_T(\omega, 0 -) = n \right], \tag{2.7}
\]
with the convention \((1 - x)^\dagger = 0\) for all \( x \in [0, 1] \).

Let us now assume \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0, 1) \). This implies in particular that the process \( X(\omega, \cdot) \) is not absorbed at \( \{0, 1\} \). As in the annealed case, one would like to use the previous moment duality to characterize the asymptotic quenched distribution of \( X \). However, in the quenched setting the situation is more involved. The reason is that, for a given realization of the environment \( \omega \), as \( T \) tends to infinity, the distribution of \( X(\omega, T) \) depends strongly on the environment near instant \( T \) (and weakly on the environment that is far away in the past). Hence, unless \( \omega \) is constant after some fixed time \( t_0 \), \( X(\omega, T) \) will not converge in distribution when \( T \) goes to infinity. In contrast, for a given realization of the environment \( \omega \) in \( (-\infty, 0] \), we will see that the distribution of \( X(\omega, 0) \), conditionally on \( X(\omega, -T) = x \), has a limit distribution, and we will characterize its law with the help of the moment duality (2.7).

For \( n \in \mathbb{N}_0 \), we define the absorption probabilities
\[
W_n(\omega) := \mathbb{P}^\omega(\exists \beta \geq 0 \text{ s.t. } R_0(\omega, \beta) = 0 \mid R_0(\omega, 0 -) = n).
\]
Clearly \( W_0(\omega) = 1 \).

**Theorem 2.12** (Quenched type-frequency from the far past). Let \( \omega \) be a simple environment in \( (-\infty, 0] \). Assume that \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0, 1) \). Then for any \( x \in (0, 1) \), the distribution of \( X(\omega, 0) \) conditionally on \( \{X(\omega, -T) = x\} \) has a limit distribution when \( T \) goes to infinity, that is a function of \( w \) and that does not depend on \( x \). Moreover, this limit distribution \( \mathcal{L}^\omega \) satisfies
\[
\int_0^1 (1 - y)^n \mathcal{L}^\omega(dy) = W_n(\omega), \quad n \in \mathbb{N}, \tag{2.8}
\]
and the convergence of moments is exponential, i.e.
\[
\mathbb{E}^{\omega} \left[ (1 - X(\omega, 0))^n \mid X(\omega, -T) = x \right] - W_n(\omega) \leq e^{-\theta_0 n T}, \quad n \in \mathbb{N}. \tag{2.9}
\]
Under the additional assumption that the selection parameter \( \sigma \) is equal to zero, we can go further and express \( W_n(\omega) \) as a function of the environment \( \omega \). This will be possible thanks to the following explicit diagonalization of the matrix \( Q^\omega_T \) (the transition matrix of the process \( R \) under the null environment).
Lemma 2.13. Assume that $\sigma = 0$ and set, for $k \in \mathbb{N}_0$, $\lambda_k := -q_k^0(k,k)$, and, for $k \in \mathbb{N}$, $\gamma_k := q_k^0(k,k-1)$. In addition, let

- $D_t$ be the diagonal matrix with diagonal entries $(-\lambda_i^t)_{i \in \mathbb{N}_0}$.
- $U_t := (u_{i,j}^t)_{i,j \in \mathbb{N}_0}$, where $u_{i,i}^t := 1$ and $u_{i,j}^t := 0$ for $j \in \mathbb{N}_0$, and, for $i \in \mathbb{N}_0$
  \[
  u_{i,j}^t := \prod_{l=j+1}^i \left( \frac{\gamma_{l+1}^t}{\lambda_i^t - \lambda_l^t} \right) \quad \text{for } j \in [i], \quad u_{i,j}^t := 0, \quad \text{for } j > i \quad \text{and} \quad u_{i,i}^t := \theta v_0 \prod_{k=1}^i \prod_{l=k+1}^i \frac{\gamma_{l+1}^t}{\lambda_l^t - \lambda_i^t},
  \] (2.10)

- $V_t := (v_{i,j}^t)_{i,j \in \mathbb{N}_0}$, where $v_{i,j}^t := 1$ and $v_{i,j}^t := 0$ for $j \in \mathbb{N}_0$, and, for $i \in \mathbb{N}$
  \[
  v_{i,j}^t := \prod_{l=j}^{i-1} \left( -\frac{\gamma_{l+1}^t}{\lambda_i^t - \lambda_l^t} \right) \quad \text{for } j \in [i], \quad v_{i,j}^t := 0, \quad \text{for } j > i \quad \text{and} \quad v_{i,i}^t := -\theta v_0 \prod_{k=1}^i \prod_{l=k}^{i-1} \left( -\frac{\gamma_{l+1}^t}{\lambda_l^t - \lambda_i^t} \right),
  \] (2.11)

with the convention that an empty sum equals 0 and an empty product equals 1. Then, we have
\[
Q_t^0 = U_t D_t V_t \quad \text{and} \quad U_t V_t = V_t U_t = \text{Id}.
\]

Now, we consider the polynomials $P_k^t$, $k \in \mathbb{N}_0$, defined via
\[
P_k^t(x) := \sum_{i=0}^k v_{i,k}^t x^i, \quad x \in [0,1]. \tag{2.12}
\]

In addition, for $z \in (0,1)$, we define the matrices $B(z) := (B_{i,j}(z))_{i,j \in \mathbb{N}_0}$ and $\beta(z) := (\beta_{i,j}(z))_{i,j \in \mathbb{N}_0}$ via
\[
B_{i,j}(z) := \begin{cases} \mathbb{P}(i + B_i(z) = j) & \text{for } i,j \in \mathbb{N}, \\ 1 & \text{for } i = j \in \{0,1\}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \beta(z) = U_t^T B(z) T V_t^T, \tag{2.13}
\]

where $B_i(z) \sim \text{Bin}(i, z)$. It will be justified in the proof of Theorem 2.14 that the matrix product defining $\beta(z)$ is well-defined.

Theorem 2.14. Assume that $\sigma = 0$, that $\theta > 0$ and $\nu_0, \nu_1 \in (0,1)$, and let $\omega$ be a simple environment. Let $N := N(T)$ be the number of jumps of $\omega$ in $(-T,0)$ and let $(T_i)_{i=0}^N$ be the sequence of those jumps in decreasing order complemented by $T_0 := 0$. For any $m \in [N]$, we define the matrix $A_m^t(\omega) := (A_{i,j}^t(\omega))_{i,j \in \mathbb{N}_0}$ via
\[
A_m^t(\omega) := \beta(\Delta \omega(T_m)) \exp \left( (T_{m-1} - T_m) D_t \right). \tag{2.14}
\]

Then, for all $x \in (0,1)$, $n \in \mathbb{N}$, we have
\[
E^n [(1 - X(\omega,0))^n | X(\omega, -T) = x] = \sum_{k=0}^{nN} C_{n,k}^t(\omega, T) P_k^t(1 - x), \tag{2.15}
\]

where the matrix $C^t(\omega, T) := (C_{n,k}^t(\omega, T))_{k,n \in \mathbb{N}_0}$ is given by
\[
C^t(\omega, T) := U_t \left[ A^t_{\omega}(\omega) A_{\omega,1}(\omega) \cdots A_{\omega,k}(\omega) \right]^T \exp \left( (T_N + T) D_t \right), \tag{2.16}
\]

with the convention that an empty product of matrices is the identity matrix. Moreover, for all $n \in \mathbb{N}$,
\[
W_n(\omega) = C_{n,0}^t(\omega, \infty) := \lim_{T \to \infty} C_{n,0}^t(\omega, T) = \lim_{T \to \infty} \left( U_t \left[ A_{\omega}(\omega) A_{\omega,1}(\omega) \cdots A_{\omega,k}(\omega) \right]^T \right)_{n,0}, \tag{2.17}
\]

where the previous limits are well-defined.

Remark 2.7. It will be justified in the proof of Theorem 2.14 that the matrix products appearing in (2.14) and (2.16) are well-defined.

Remark 2.8. Note that if $\omega$ has no jumps in $(-T,0)$, then $N(T) = 0$ and $T_N(T) = T_0 = 0$. Hence, Eq. (2.16) reads $C^t(\omega, T) = U_t \exp(T D_t)$. In particular, for the null-environment, we obtain $W_n(0) = u_{n,0}^t$. 

Let us consider a realization $G$ where the focus was on the study of the distribution of the type frequency from a backward perspective, in this section we are interested in the ancestral type distribution. We start by making this notion precise.

Now, consider a simple environment $V$ let $j$ respectively; a line having level $i$ incoming (resp. continuing) line of the descendant line with level $j > i$ to the lines in $V$.

Given $G_T$ and $c$, we propagate types (forward in time) along the lines of $G_T$ keeping track, at any time $T$, of the true ancestor in $V_T$ of each line in $V_T$. We denote by $a_c(G_T)$ the type of the ancestor in $V_T$ of the single line in $V_0$. Assume that, under $P_x$, $c$ assigns independently to each line type 0 with probability $x$ and type 1 with probability $1 - x$. We call annealed ancestral type distribution at time $T$ to

$$h_T(x) := P_x(a_c(G_T) = 0), \quad x \in [0, 1].$$

Now, consider a simple environment $\omega$ and let $G_T(\omega) := (G_T(\omega, \beta))_{\beta \in [0, T]}$ be the corresponding quenched ASG in $[0, T]$ started with one line, representing the individual sampled at forward time $T$. For $\beta \in [0, T]$, let $V_\beta(\omega)$ be the set of lines present at time $\beta$ in $G_T(\omega)$. For a given type assignment $c : V_T(\omega) \to \{0, 1\}$ to the lines in $V_T(\omega)$, we proceed as in the annealed case and denote by $a_c(G_T(\omega))$ the type of the ancestor in $V_T(\omega)$ of the single line in $V_0(\omega)$. We call quenched ancestral type distribution at time $T$ to

$$h_T^\omega(x) := P_x^\omega(a_c(G_T(\omega)) = 0), \quad x \in [0, 1],$$

where, under $P_x^\omega$, $c$ assigns independently to each line in $V_T(\omega)$ type 0 with probability $x$ and type 1 with probability $1 - x$.

In the case of the null environment, the ancestral type distribution can be expressed in terms of the block-counting process of the pruned lookdown ASG (pruned LD-ASG), see [28]. The main goals of this section are: a) to incorporate the effect of the environment in the construction of the pruned LD-ASG, b) to express the annealed and quenched ancestral type distributions in terms the corresponding block-counting processes, and c) to infer their asymptotic behavior as $T$ tends to infinity.

2.6. The pruned LD-ASG. In this subsection, we adapt the construction of the pruned LD-ASG to incorporate the effect of the environment. We first build the lookdown-ASG (LD-ASG), which consists in attaching levels to each line in the ASG encoding the hierarchy given by the pecking order. This is done as follows. Consider a realization of the (annealed or quenched) ASG in $[0, T]$ starting with one line. The starting line gets level 1. When two lines, one with level $i$ and the other with level $j > i$, coalesce, the resulting line is assigned level $i$; the level of each line with level $k + j$ before the coalescence is decreased by 1. When a group of lines with levels $i_1 < i_2 < \ldots < i_N$ experiences a simultaneous branching, the incoming (resp. continuing) line of the descendant line with level $i_k$ gets level $i_k + k - 1$ (resp. $i_k + k$), respectively; a line having level $j$ before the branching, with $i_k < j < i_{k+1}$ gets level $j + k$; a line having level $j > i_N$ before the branching gets level $j + N$. Mutations do not affect the levels. See Fig. 4(left) for an illustration.

On the basis of the LD-ASG the pruned LD-ASG is obtained via an appropriate pruning of its lines. In order to describe this pruning procedure, we need first to identify a special line in the LD-ASG: the immune line. The immune line at time $\beta$ is the line in the ASG present at time $\beta$ that is the ancestor of the starting line if all the lines at time $\beta$ are assigned the unfit type. In the absence of mutations, the immune line evolves according to the following rules:

- It only changes at coalescence or branching times involving the current immune line.
- At a coalescence event involving the immune line, the merged line becomes the immune line.
- If the immune line is subject to a branching, then the continuing line becomes the immune line.

In the presence of mutations, the pruned LD-ASG is constructed simultaneously with the immune line as follows. Let $\beta_1 < \cdots < \beta_n$ be the times at which mutations occur in the LD-ASG in $[0, T]$. In the time interval $[\beta_{i-1}, \beta_i)$ the pruned LD-ASG coincides with the LD-ASG and the immune line evolves as described above. Assume now that we have constructed the pruned LD-ASG together with its immune line up to time $\beta_i - \epsilon$, where the pruned LD-ASG contains $N$ lines and the immune line has level $k_0 \in [N]$. The pruned LD-ASG is extended up to time $\beta_i$ according to the following rules:
The idea is to treat simultaneous branching events locally as simple branching events. The proof of Theorem 4 in [28] can be adapted in a straightforward way to prove Lemma 2.15. The basic definitions for the corresponding block-counting processes.

2.6.2. Ancestral type distribution: the annealed case.

We can also provide definitions for the annealed and quenched pruned LD-ASG analogous to Definitions 2.6 and 2.8. However, thanks to Lemma 2.15, in order to compute the ancestral type distribution, it is enough to keep track of the number of lines in the pruned LD-ASG. Hence, we will directly provide definitions for the corresponding block-counting processes.

Lemma 2.15 (Theorem 4 in [28]). If we assign types at instant 0 in the pruned LD-ASG, the true ancestor is the line of type 0 with smallest level or, if all lines have type 1, it is the immune line.

The proof of Theorem 4 in [28] can be adapted in a straightforward way to prove Lemma 2.15. The basic idea is to treat simultaneous branching events locally as simple branching events. So far we have defined the pruned LD-ASG in a static way (as a function of a realization of the ASG). We can also provide definitions for the annealed and quenched pruned LD-ASG analogous to Definitions 2.7 and 2.8. However, thanks to Lemma 2.15, in order to compute the ancestral type distribution, it is enough to keep track of the number of lines in the pruned LD-ASG. Hence, we will directly provide definitions for the corresponding block-counting processes.

\[q^\mu(i, j) := \begin{cases} 
i(i - 1) + (i - 1)\theta \nu_1 + \theta \nu_0 & \text{if } j = i - 1, \\
i(\sigma + \sigma_{i, 1}) & \text{if } j = i + 1, \\
\binom{i}{j} \sigma_{i, k} & \text{if } j = i + k, i \geq k \geq 2, \\
\theta \nu_0 & \text{if } 1 \leq j < i - 1, \\
-(i - 1)(i + \theta) - i\sigma - \int_{(0, 1)}(1 - (1 - y)^i)\mu(dy) & \text{if } j = i, \\
0 & \text{otherwise.} \]
The next result provides an important feature of the process $L$.

**Lemma 2.16 (Positive recurrence).** The process $L$ is positive recurrent.

As a consequence, $L$ admits a unique stationary distribution that we denote by $\pi_L$. We let $L_\infty$ be a random variable with law $\pi_L$. The following result provides the formal relation between the pruned LD-ASG and the ancestral type distribution.

**Proposition 2.17.** For all $T \geq 0$ and $x \in [0,1]$,

$$h_T(x) = 1 - \mathbb{E}[(1 - x)^L(T) \mid L(0) = 1].$$

(2.19)

Moreover, $h(x) := \lim_{T \to \infty} h_T(x)$ is well-defined, and

$$h(x) = \sum_{n \geq 0} x(1 - x)^n \mathbb{P}(L_\infty > n).$$

(2.20)

In the absence of mutations, the previous proposition leads to the following result.

**Corollary 2.18.** If $\theta = 0$, then, for any $T > 0$ and $x \in [0,1]$,

$$h_T(x) = \mathbb{E}[X(T) \mid X(0) = x].$$

Moreover, conditional on $X(0) = x$, the limit of $X(T)$ when $T \to \infty$ is almost surely well-defined and is a Bernoulli random variable with parameter $h(x)$. In particular, the absorbing states 0 and 1 are both accessible from any $x \in (0,1)$.

**Remark 2.9.** We point out that the accessibility of both boundaries given in the previous result is not covered by the criteria given in [20, Thm. 3.2]. The latter holds for a fairly general class of neutral reproduction mechanisms, but excludes the possibility of a diffusive term.

The next result characterizes the tail probabilities $a_n := \mathbb{P}(L_\infty > n)$, $n \in \mathbb{N}_0$.

**Theorem 2.19 (Fearnhead’s type recursion).** For all $n \geq 1$, we have

$$(\sigma + \theta + n + 1) a_n = \sigma a_{n-1} + (\theta \nu_1 + n + 1) a_{n+1} + \frac{1}{n} \sum_{j=1}^{n} \gamma_{n+1,j} (a_{j-1} - a_j), \quad n \in \mathbb{N},$$

(2.21)

where $\gamma_{n,j} := \sum_{k=i-j}^{k} \binom{i}{k} \sigma_{j,k}$.

2.6.3. **Ancestral type distribution: the quenched case.** In this section the environment $\omega$ on $(-\infty, \infty)$ is fixed. We assume that $\omega$ is simple, has infinitely many jumps on $[0, \infty)$ and that the distance between the successive jumps does not converge to 0. These assumptions are almost surely satisfied when $\omega$ is given by a realization of the process $J$ described in Subsection 2.2 in the case where the measure $\mu$ is finite. We first define the pruned LD-ASG in the quenched setting. Next, we provide an expression for $h^*_{i\omega}(x)$ and study its limit when $T \to \infty$.

Let us fix $T \in \mathbb{R}$ and let $(L_T(\omega, \beta))_{\beta \geq 0}$ be the block-counting process of the pruned LD-ASG $\mathcal{G}^T(\omega, \cdot)$. From construction, $(L_T(\omega, \beta))_{\beta \geq 0}$, is a continuous-time (in-homogeneous) Markov chain with values on $\mathbb{N}$. Between the jumps of the environment $\omega$, $(L_T(\omega, \beta))_{\beta \geq 0}$ has transitions rates $(q^0(i,j))_{i,j \in \mathbb{N}}$ (i.e $\mu = 0$ in (2.18)). In addition, at each time $\beta \geq 0$ such that $\Delta \omega(T - \beta) > 0$, we have, for all $i \in \mathbb{N}$ and $k \in [i]_0$,

$$\mathbb{P}^i(L_T(\omega, \beta) = i + k \mid L_T(\omega, \beta-) = i) = \binom{i}{k} (\Delta \omega(T - \beta))^k (1 - \Delta \omega(T - \beta))^{i-k}.$$

In other words, conditionally on $\{L_T(\omega, \beta-) = i\}$, we have $L_T(\omega, \beta) \sim i + Y$, where $Y \sim \text{Bin}(i, \Delta \omega(s))$. The relation between the pruned LD-ASG and the ancestral type distribution can be stated as follows.

**Proposition 2.20.** For all $T > 0$, $x \in [0,1]$, $n \in \mathbb{N}$, we have

$$h^*_{\omega}(x) = 1 - \mathbb{E}^x[(1 - x)^L(\omega, T-) \mid L_T(\omega, 0-) = 1].$$

(2.22)
Remark 2.10. In the case $\theta = 0$, the processes $(R_T(\omega, \beta))_{\beta \geq 0}$ and $(L_T(\omega, \beta))_{\beta \geq 0}$ both coincide with the block-counting process of the quenched ASG. In particular, combining Theorem 2.11 together with Proposition 2.20 leads to $h_T^\omega(x) = \mathbb{E}[X(\omega, T) \mid X(\omega, 0) = x]$. Moreover, the diffusion $X(\omega, \cdot)$ is eventually almost surely absorbed at $\{0, 1\}$. In particular, $h^\omega(x)$, the limit of $h_T^\omega(x)$ as $T \to \infty$, exists and equals the probability of fixation of the fit type.

In contrast to the annealed case, in the quenched case the block-counting process of the pruned LD-ASG does not have a stationary distribution. However, for the limit of $h_T^\omega(x)$ when $T \to \infty$ to be well-defined, we only need that the distribution of $L_T(\omega, T\cdot)$ admits a limit when $T \to \infty$. The next Theorem provides such a convergence result.

**Theorem 2.21.** For any $n \geq 1$, the distribution of $L_T(\omega, T\cdot)$ conditionally on $\{L_T(\omega, 0\cdot) = n\}$ has a limit distribution $\mu^\omega \in \mathcal{M}_1(\mathbb{N})$, when $T \to \infty$, that does not depend on $n$. In particular, the limit $h^\omega(x) := \lim_{T \to \infty} h_T^\omega(x)$ is well-defined and

$$h^\omega(x) := 1 - \sum_{n=1}^{\infty} \mu^\omega(\{n\})(1 - x)^n.$$ (2.23)

Moreover, if $\theta \nu_0 > 0$, then for any $x \in [0, 1]$ and $t > 0$ we have

$$|h^\omega(x) - h_T^\omega(x)| \leq 2e^{-\theta \nu_0 t}.$$ (2.24)

Under the additional assumption that the selection parameter $\sigma$ equals to zero, we can go further and express $h_T^\omega(x)$ as a function of the environment $\omega$. The next result will be crucial in order to obtain the expression for $h_T^\omega(x)$.

**Lemma 2.22.** Assume that $\sigma = 0$ and set, for $k \in \mathbb{N}$, $\lambda_k := -q^0(k, k)$, and, for $k \in \mathbb{N}$, $\gamma_k := q^0(k, k-1)$. In addition, let

- $D$ be the diagonal matrix with diagonal entries $(-\lambda_i)_{i \in \mathbb{N}}$.
- $U := (u_{i,j})_{i,j \in \mathbb{N}}$ where, for all $i \in \mathbb{N}$, $u_{i,1} = 1$, $u_{i,j} = 0$ for $j > i$ and, when $i \geq 2$, $u_{i,i-1} := \gamma_i/(\lambda_i - \lambda_{i-1})$ and the coefficients $(u_{i,j})_{j \in [i-2]}$ are defined via the recurrence relation

$$u_{i,j} := \frac{1}{\lambda_i - \lambda_j} \left( \gamma_i u_{j-1}^i + \nu_0 \theta \sum_{l=j}^{i-2} u_{l,j}^i \right).$$ (2.25)

- $V := (v_{i,j})_{i,j \in \mathbb{N}}$ where, for all $i \in \mathbb{N}$, $v_{i,1} = 1$, $v_{i,j} = 0$ for $j > i$ and, when $i \geq 2$, the coefficients $(v_{i,j})_{j \in [i-1]}$ are defined via the recurrence relation

$$v_{i,j} := \frac{-1}{(\lambda_i - \lambda_j)} \left( \sum_{l=j+2}^{i} v_{l,j} \right) \nu_0 \theta + v_{i,j+1} \gamma_{j+1}.$$ (2.26)

with the convention that an empty sum equals 0. Then, we have

$$Q = UV^\top \quad \text{and} \quad UV = VU = Id.$$

**Remark 2.11.** Lemma 2.22 can be proved the same way as Lemma 2.13 so its proof will be omitted.

Now, we consider the polynomials $P_k, k \in \mathbb{N}$ defined via

$$P_k(x) := \sum_{i=0}^{k} v_{k,i} x^i.$$ (2.27)

In addition, for $z \in (0, 1)$, we define the matrices $\mathcal{B}(z) := (B_{i,j}(z))_{i,j \in \mathbb{N}}$ and $\beta(z) := (\beta_{i,j}(z))_{i,j \in \mathbb{N}}$ via

$$B_{i,j}(z) := P(i + B_i(z) = j), \quad i, j \in \mathbb{N} \quad \text{and} \quad \beta(z) = U^\top \mathcal{B}(z) \mathcal{V}^\top,$$ (2.28)

where $B_i(z) \sim \text{Bin}(i, z)$. The fact that the matrix product defining $\beta(z)$ is well-defined can be justified similarly as in the proof of Theorem 2.13 where we show that $\beta^l(z)$, defined in (2.13), is well-defined.
Theorem 2.23. Assume that $\sigma = 0$ and let $\omega$ be a simple environment with infinitely many jumps on $[0, \infty)$ and such that the distance between the successive jumps does not converge to 0. Let $N$ be the number of jumps of $\omega$ in $(0, T)$ and let $(T_i)_{i=1}^N$ be the sequence of those jumps in increasing order. We set $T_0 := 0$ for convenience. For any $m \in [N]$, we define the matrix $A_m(\omega) := (A_{i,j}^m(\omega))_{i,j \in \mathbb{N}}$ by

$$A_m(\omega) := \exp((T_m - T_{m-1})D) \beta(\Delta \omega(T_m)).$$

(2.29)

Then for all $x \in (0, 1)$, $n \in \mathbb{N}$, we have

$$h_n^x(x) = 1 - \sum_{k=1}^{2^N} C_{1,k}(\omega, T) P_k(1 - x),$$

(2.30)

where the matrix $C(\omega, T) := (C_{n,k}(\omega, T))_{k,n \in \mathbb{N}}$ is given by

$$C(\omega, T) := U \exp((T - T_N)D) [A_1(\omega)A_2(\omega) \cdots A_N(\omega)]^\top.$$  

(2.31)

Moreover, for any $x \in (0, 1)$,

$$h^\omega(x) = 1 - \sum_{k=1}^{\infty} C_{1,k}(\omega, \infty) P_k(1 - x),$$

(2.32)

where the series in (2.32) is convergent and where

$$C_{1,k}(\omega, \infty) := \lim_{m \to +\infty} \left(U [A_1(\omega)A_2(\omega) \cdots A_m(\omega)]^\top\right)_{1,k},$$

(2.33)

and the above limit is well-defined.

Remark 2.12. The fact that the matrix products (2.29) and (2.31) are well-defined can be justified similarly as in the proof of Theorem 2.14 where we show that the matrix products (2.14) and (2.16) are well-defined.

3. Moran models: continuity with respect to the environment

In this section, we aim to prove Theorem 2.11 which states, in the context of a Moran model population, the continuity of the type-frequency process with respect to the environment. On the one hand, the paths of the type-frequency process are considered as elements of $\mathbb{D}_T$, which is endowed with the Skorokhod topology, i.e. the topology induced by the metric $d^\ast_T$ defined in Appendix A. On the other hand, the environment is given by means of a function in the set

$$D^\ast_T := \{\omega \in \mathbb{D}_T : \omega(0) = 0, \Delta \omega(t) \in [0, 1) \text{ for all } t \in [0, T], \omega \text{ is non-decreasing and pure-jump}\},$$

which is endowed with the topology induced by the metric $d^\ast_T$ defined in Appendix A.

Let us denote by $\mu_N(\omega)$ the law of $(X_N(\omega, t))_{t \in [0, T]}$. Theorem 2.11 hence states the continuity of the mapping $\omega \mapsto \mu_N(\omega)$, where the set of probability measures on $\mathbb{D}_T$ is equipped with the topology of weak convergence of measures. We will use the fact that the topology of weak convergence of probability measures on a complete and separable metric space $(E, d)$ is induced by the metric $g_E$ defined in Appendix B. We will also prove some results about uniform convergence of finite dimensional distributions. For this we introduce some notation. For $\omega \in \mathbb{D}_T^n$, $n \in \mathbb{N}$ and $\bar{t} := (t_i)_{i \in [n]} \in [0, T]^n$, $\mu^\bar{t}(\omega)$ denotes the law of $(X_N(\omega, t_i))_{i \in [n]}$. Here, we consider $[0, 1]^n$ equipped with the distance $d_1$ defined via $d_1((x_i)_{i \in [n]}, (y_i)_{i \in [n]}):= \sum_{i \in [n]} |x_i - y_i|$.

We start with a result that will be useful in order to get rid of the small jumps of the environment. For $\delta > 0$ and $\omega \in \mathbb{D}_T^n$, we define $\omega^\delta, \omega_\delta \in \mathbb{D}_T$ via

$$\omega^\delta(t) := \sum_{u \in [0,t] : \Delta \omega(u) \geq \delta} \Delta \omega(u) \text{ and } \omega_\delta(t) := \sum_{u \in [0,t] : \Delta \omega(u) < \delta} \Delta \omega(u).$$

Clearly, $\omega^\delta$ is simple and $\omega = \omega^\delta + \omega_\delta$. Moreover, $\omega_\delta \to 0$ pointwise as $\delta \to 0$, and hence for any $t \in [0, T]$,

$$d^\ast_T(\omega, \omega^\delta) \leq \sum_{u \in [0,T]} |\Delta \omega(u) - \Delta \omega^\delta(u)| = \omega_\delta(t) \xrightarrow{\delta \to 0} 0.$$
Proposition 3.1. Let \( \omega \in \mathbb{D}_T^r, N \geq 1, T > 0 \). Assume that for any \( \delta > 0 \), we have \( X_N(\omega^\delta, 0) = X_N(\omega, 0) \), then
\[
\varrho_{[0,T]}(\mu^\delta_N(\omega^\delta), \mu^\delta_N(\omega)) \leq n \omega(\omega^\delta) e^{\sigma t + \omega(\omega^\delta)}, \quad \forall t \in [0, T]^n, t \in \mathbb{N},
\]
where \( t_* := \max_{t \in [n]} t_i \). Moreover,
\[
\varrho_{[0,T]}(\mu_N(\omega^\delta), \mu_N(\omega)) \leq \omega(T) e^{(1+\sigma)t \omega(T)}.
\]
In particular,
\[
(X_N(\omega^\delta, t))_{t \in [0, T]} \overset{d}{=} (X_N(\omega, t))_{t \in [0, T]}.
\]

Proof. For \( \delta > 0 \), we couple in \([0, T]\) a Moran model with parameters \( (s, u, \nu_0, \nu_1) \) and environment \( \omega \) to a Moran model with parameters \( (s, u, \nu_0, \nu_1) \) and environment \( \omega^\delta \) (both of size \( N \)) by using: (1) the same initial type configuration, (2) the same basic background, and (3) the same environmental background (see Section 2.1.2 for the definitions of basic and environmental backgrounds).

For any \( t \in [0, T] \) and \( a, b \in \{0, 1\} \), we denote by \( X_{N, t}^{a,b}(t) \) the proportion of individuals that at time \( t \) have type \( a \) under the environment \( \omega \) and type \( b \) under the environment \( \omega^\delta \). Clearly, we have
\[
|X_{N}(\omega^\delta, t) - X_N(\omega, t)| = |X^1_{N, t}(t) - X^0_{N, t}(t)| \leq X^1_{N, t}(t) + X^0_{N, t}(t) := Z_N(t).
\]
Note that \( Z_N(t) \) corresponds to the proportion of individuals that have a different type at time \( t \) under \( \omega \) and \( \omega^\delta \). Let us assume that at time \( t \) a graphical element arises in the basic background. If the graphical element corresponds to a mutation event, then \( Z_N(t) \leq Z_N(t^-) \). If the graphical element is a neutral reproduction, we have
\[
\mathbb{E}[Z_N(t) \mid Z_N(t^-)] = Z_N(t^-) + \frac{1}{N} Z_N(t^-)(1 - Z_N(t^-)) - \frac{1}{N} (1 - Z_N(t^-)) Z_N(t^-) = Z_N(t^-).
\]
If the graphical element corresponds to a selective event, then
\[
\mathbb{E}[Z_N(t) \mid Z_N(t^-)] \leq \left(1 + \frac{1}{N}\right) Z_N(t^-).
\]
Now, let us assume that \( 0 \leq s < t \leq T \) are such that the interval \((s, t)\) does not contain jumps of \( \omega^\delta \) nor selective events. In particular, in this interval only the population driven by \( \omega \) is affected by the environment. Moreover, since neutral and mutation events do not increase the expected value of \( Z_N(u), u \in [0, T] \), we obtain
\[
\mathbb{E}[Z_N(t^-) \mid Z_N(s)] \leq Z_N(s) + \sum_{u \in [s, t]} \Delta \omega(u).
\]
In addition, if at time \( t \) the environment \( \omega^\delta \) jumps (there are only finitely many of these jumps), then
\[
\mathbb{E}[Z_N(t) \mid Z_N(t^-)] = Z_N(t^-)(1 + \Delta \omega(t)).
\]
Let \( 0 \leq t_1 < \cdots < t_n \leq T \) be the jumping times of \( \omega^\delta \). From the previous discussion, we obtain
\[
\mathbb{E}[Z_N(t_{i+1}) \mid Z_N(t_i)] \leq \mathbb{E} \left[ \left(1 + \frac{1}{N}\right)^{K_i} \right](Z_N(t_i) + \epsilon_i(\delta))(1 + \Delta \omega(t_{i+1})) \tag{3.3}
\]
where \( \epsilon_i(\delta) := \sum_{u \in (t_i, t_{i+1})} \Delta \omega(u) \) and \( K_i \) is the number of selective events in \((t_i, t_{i+1})\). Note that \( K_i \) has a Poisson distribution with parameter \( N \sigma(t_{i+1} - t_i) \), and hence
\[
\mathbb{E}[Z_N(t_{i+1}) \mid Z_N(t_i)] \leq e^{\sigma(t_{i+1} - t_i)}(Z_N(t_i) + \epsilon_i(\delta))(1 + \Delta \omega(t_{i+1})).
\]
Iterating this formula and using that \( Z_N(0) = 0 \) yields
\[
\mathbb{E}[Z_N(t)] \leq e^{\sigma \omega(t)} \prod_{i : t_i \leq s} (1 + \Delta \omega(t_i)) \leq \omega(T) e^{\sigma t + \sum_{u \in [0,t]} \Delta \omega(u)}, \tag{3.4}
\]
Recall the definition of the space \( BL(E) \) from Appendix [13]. Note that for any \( n \geq 1 \) and \( F \in BL([0, 1]^n) \) with \( \|F\|_{BL} \leq 1 \) we have
\[
\left| \int Fd\mu_N(\omega^\delta) - \int Fd\mu_N(\omega) \right| = \left| \mathbb{E} \left[ F((X_N(\omega^\delta, t_j))_{j \in [n]}) \right] - \mathbb{E} \left[ F((X_N(\omega, t_j))_{j \in [n]}) \right] \right|
\]
and if we choose the above coupling for \( X_N(\omega^t, t) \) and \( X_N(\omega, t) \) we get that the above equals

\[
\mathbb{E} \left[ F\left( (X_N(\omega^t, t_j))_{j \in [n]} \right) - F\left( (X_N(\omega, t_j))_{j \in [n]} \right) \right] \leq \mathbb{E} \left[ d_1 \left( (X_N(\omega^t, t_j))_{j \in [n]} , (X_N(\omega, t_j))_{j \in [n]} \right) \right]
\]

\[
= \mathbb{E} \left[ \sum_{j \in [n]} |Z_N(t_j)| \right] \leq \sum_{j \in [n]} \omega_k(t_j)e^{\sigma t_j + \sum_{u \in [0,t_j]} \Delta \omega(u)} ,
\]

where the last bound comes from (3.4) applied at each \( t_j \). Taking the supremum over all \( F \in \text{BL}([0,1]^n) \) with \( \| F \|_{\text{BL}} \leq 1 \) and using the definition of the distance \( d_1(\omega^t, \omega) \) in Appendix B we get (3.1). Now, define \( Z_N(t) := \sup_{u \in [0,t]} Z_N(u) \). If at time \( t \) a neutral event occurs, then

\[
\mathbb{E}[Z_N(t) | Z_N(t-) \leq \left( 1 + \frac{1}{N} \right) Z_N(t-).
\]

Other events can be treated as before, leading to (3.3) with \( K \), being this the number of selective and neutral events in \( (t_i, t_{i+1}) \). Hence, Eq. (3.2) follows similarly as (3.1). The convergence of \( X_N(\omega^t, \cdot) \) towards \( X_N(\omega, \cdot) \) is a direct consequence of (3.2).

\[\square\]

**Proposition 3.2.** Let \( \omega \in \mathbb{D}_T^* \) and \( \{\omega_k\}_{k \in \mathbb{N}} \subset \mathbb{D}_T^* \) be such that \( d_T(\omega_k, \omega) \rightarrow 0 \) as \( k \rightarrow \infty \). If \( \omega \) is simple and, for any \( k \in \mathbb{N} \), \( X_N(\omega_k, 0) = X_N(\omega_k, 0) \), then

\[
(X_N(\omega_k, t))_{t \in [0,T]} \xrightarrow{d} (X_N(\omega, t))_{t \in [0,T]} .
\]

**Proof.** The proof consists of two parts. In the first part, we construct a time deformation \( \lambda_k \in C_T^f \) with suitable properties. In the second part, we compare \( X_N(\omega_k, \lambda_k(\cdot)) \) and \( X_N(\omega, \cdot) \) under an appropriate coupling of the underlying Moran models.

**Part 1:** Without loss of generality, we can assume that \( d_T(\omega_k, \omega) > 0 \) for all \( k \in \mathbb{N} \). Set \( \epsilon_k := 2d_T(\omega_k, \omega) \), so that \( d_T(\omega_k, \omega) < \epsilon_k \). From definition of the metric \( d_T \) in Appendix A there is \( \varphi_k \in C_T^f \) such that

\[
\| \varphi_k \|_T^0 \leq \epsilon_k \text{ and } \sum_{u \in [0,T]} |\Delta \omega(u) - \Delta (\omega_k \circ \varphi_k)(u)| \leq \epsilon_k .
\]

Denote by \( r_1 < \cdots < r_n \) the consecutive jumps of \( \omega \) in \( [0,T] \). For simplicity we assume that \( 0 < r_1 \leq r_n < T \), but the proof can be easily adapted to the case where \( \omega \) jumps at \( T \). Set \( \gamma_k := T \sqrt{\epsilon_k} - 1 \). In the remainder of the proof we assume that \( k \) is sufficiently large, such that \( \gamma_k \leq \min_{i \in \mathbb{N}} (r_{i+1} - r_i)/3 \), where \( r_0 := 0 \) and \( r_{n+1} := T \). This condition ensures that the intervals \( I_i := [r_i - \gamma_k, r_i + \gamma_k] \), \( i \in \mathbb{N} \), are disjoint and contained in \([0, T] \). Now, we define \( \lambda_k : [0,T] \rightarrow [0,T] \) via

- For \( u \notin \cup_{i=0}^{n} I_i \) : \( \lambda_k(u) := u \).
- For \( u \in [r_i - \gamma_k, r_i) \) : \( \lambda_k(u) := \varphi_k(r_i) + m_i(u - r_i) \), where \( m_i := (\varphi_k(r_i) - r_i + \gamma_k)/\gamma_k \).
- For \( u \in (r_i, r_i + \gamma_k) \) : \( \lambda_k(u) := \varphi_k(r_i) + \bar{m}_i(u - r_i) \), where \( \bar{m}_i := (r_i + \gamma_k - \varphi_k(r_i))/\gamma_k \).

For \( k \) large enough so that \( \epsilon_k < \log 2 \), we can infer from \( \| \varphi_k \|_T^0 \leq \epsilon_k \) and from \( \gamma_k = T \sqrt{\epsilon_k} - 1 \) that \( m_i \) and \( \bar{m}_i \) are positive. It is then straightforward to check that \( \lambda_k \) has the following properties:

- \( \lambda_k \in C_T^f \)
- \( \lambda_k(I_i^0) = I_i \), \( i \in \mathbb{N} \).
- \( \sum_{u \in [0,T]} |\Delta \omega(u) - \Delta \lambda_k(u)| \leq \epsilon_k \), where \( \lambda_k := \omega \circ \lambda_k \).

Moreover, since \( \| \varphi_k \|_T^0 \leq \epsilon_k \), we infer that \( \varphi_k(r_i) \in [\epsilon_k r_i, \epsilon_k r_i] \). It follows that, for \( k \) large enough, we have

\[
1 - 2\sqrt{\epsilon_k} - 1 \leq m_i \leq 1 + 2\sqrt{\epsilon_k} - 1 , \quad i \in \mathbb{N}
\]

and the same holds for \( \bar{m}_i \). Therefore, the right derivative of \( \lambda_k(\cdot) \), \( \lambda_k'(t) \), belongs to \([1 - 2\sqrt{\epsilon_k} - 1, 1 + 2\sqrt{\epsilon_k} - 1]\) for \( t \in [0,T] \). We thus have that for any \( s, t \in [0,T] \) with \( s \neq t \), \( (\lambda_k(s) - \lambda_k(t))/(s-t) \) belongs to \([1 - 2\sqrt{\epsilon_k} - 1, 1 + 2\sqrt{\epsilon_k} - 1]\). We thus get that for \( k \) large enough

\[
\frac{\lambda_k(s) - \lambda_k(t)}{s - t} , \quad \frac{s - t}{\lambda_k(s) - \lambda_k(t)} \leq 1 + 3\sqrt{\epsilon_k} - 1 , \quad i \in \mathbb{N}.
\]

Hence, using that \( \log(1 + x) \leq x \) for \( x > -1 \), leads to \( \| \lambda_k \|_T^0 \leq 3\sqrt{\epsilon_k} - 1 \) for large \( k \).
Part 2: Recall the definition of the basic background and of the environmental background in Subsection 2.1. For $\delta > 0$, we couple in $[0, T]$ a Moran model with parameters $(s, u, v_0, v_1)$ and environment $\omega$ to a Moran model with parameters $(s, u, v_0, v_1)$ and environment $\omega_k$ (both of size $N$) by using: (1) the same initial type configuration, (2) the same basic background, and (3) using in the second population the environmental background of the first one time-changed by $\lambda_k^{-1}$. By construction of the function $\lambda_k$, under this coupling, the Moran model associated to $\omega$ and the Moran model associated to $\omega_k$, time-changed by $\lambda_k$, experience the same basic events out of the time-intervals $I^k_t$, and at times $r_i$, the success of simultaneous environmental reproductions is decided according to the same uniform random variables. For any $t \in [0, T]$, we denote by $Z_N(t)$ the proportion of individuals that have a different type at time $t$ for $\omega$ and at time $\lambda_k(t)$ for $\omega_k$. Moreover, we set $Z_N^+(t) := \sup_{u \in [0, t]} Z_N(u)$.

Consider the event $E_k := \{\text{there are no basic events in } \cup_{i \in [n]} I^k_t\}$. Note that

$$P(E_k^c) \leq n \left( 1 - e^{-2N(1+\sigma+\theta)\gamma_k} \right).$$

Moreover, on the event $E_k$, only the population driven by $\omega_k$ can change in $(r_i, r_i + \gamma_k]$, and this can only be due to environmental events. Hence,

$$E[Z_N^+(r_i + \gamma_k) 1_{E_k}] \leq E[Z_N^+(r_i) 1_{E_k}] + \sum_{u \in (r_i, r_i + \gamma_k]} \Delta \omega_k(u). \quad (3.6)$$

A similar argument, allows to infer that

$$E[Z_N^+(r_{i+1} - 1_{E_k})] \leq E[Z_N^+(r_{i+1} - \gamma_k) 1_{E_k}] + \sum_{u \in (r_{i+1} - \gamma_k, r_{i+1})} \Delta \omega_k(u). \quad (3.7)$$

Moreover, since in the interval $(r_i + \gamma_k, r_{i+1} - \gamma_k]$ there are no simultaneous jumps of the two environments, we can proceed as in the proof of Proposition 3.1 to obtain

$$E[Z_N^+(r_{i+1} - \gamma_k) 1_{E_k}] \leq e^{(1+\sigma)(r_{i+1} - r_i)} \left( E[Z_N^+(r_i + \gamma_k) 1_{E_k}] + \sum_{u \in (r_i + \gamma_k, r_{i+1} - \gamma_k]} \Delta \omega_k(u) \right). \quad (3.8)$$

Moreover, at time $r_{i+1}$, there are two possible contributions to take into account: (i) the contribution of selective arrows arising simultaneously in both environments, and (ii) the contribution of selective arrows arising only on the environment with the biggest jump. This leads to

$$E[Z_N^+(r_{i+1}) 1_{E_k}] \leq E[Z_N^+((r_{i+1}) - 1_{E_k})]\left(1 + \Delta \omega(r_{i+1}) \wedge \Delta \omega_k(r_{i+1}) + |\Delta \omega(r_{i+1}) - \Delta \omega_k(r_{i+1})| \right]. \quad (3.9)$$

Using (3.6), (3.7), (3.8) and (3.9), we obtain

$$E[Z_N^+(r_{i+1}) 1_{E_k}] \leq e^{(1+\sigma)(r_{i+1} - r_i)} \left( E[Z_N^+(r_i) 1_{E_k}] + \sum_{u \in (r_i, r_{i+1})} |\Delta \omega(u) - \Delta \omega_k(u)| (1 + \Delta \omega(r_{i+1})) \right).$$

Iterating this inequality, using that $Z_N^+(0) = 0$, and adding the contribution of the interval $(r_n + \gamma_k, T]$, we obtain

$$E[Z_N^+(T) 1_{E_k}] \leq \epsilon_k e^{(1+\sigma)T + \sum_{u \in (0, T]} \Delta \omega(u)}.$$ 

Summarizing, we get

$$E \left[ d_{T}^0(X_N(\omega, \cdot), X_N(\omega_k, \cdot)) \right] \leq 2n \left( 1 - e^{-2N(1+\sigma+\theta)\gamma_k} \right) + \sqrt{\epsilon_k} - 1 \vee \left( \epsilon_k e^{(1+\sigma)T + \sum_{u \in (0, T]} \Delta \omega(u)} \right).$$

The result follows. \hfill \Box

Proof of Theorem 2.7 (Continuity). If $\omega$ has a finite number of jumps, the result follows directly from Proposition 3.2. In the general case, note that for any $\delta > 0$, we have

$$g_{\delta, T}(MN(\omega_k), MN(\omega)) \leq g_{\delta, T}(MN(\omega), MN(\omega_k)) + g_{\delta, T}(MN(\omega_k), MN(\omega)) + g_{\delta, T}(MN(\omega), MN(\omega)),$$

where $\omega_k^\delta(t) := \sum_{u \in [0, t], \Delta \omega_k(u) \geq \delta} \Delta \omega_k(u)$.

Now, we claim that, for any $\delta \in A_\omega := \{d > 0 : \Delta \omega(u) \neq d \text{ for any } u \in [0, T]\}$, we have

$$d_{T}^\delta(\omega_k, \omega^\delta) \underset{k \to \infty}{\longrightarrow} 0. \quad (3.11)$$
Assume the claim is true and let $\delta \in A_{\omega}$. Note that for any $\lambda \in \mathcal{C}_{T}^{\uparrow}$, we have

$$\omega_{k,\delta}(T) := \sum_{u \in [0,T] : \Delta \omega_{k}(u) < \delta} \Delta \omega_{k}(u) = \omega_{k}(T) - \omega_{k}^{\delta}(T) = \omega_{k}(\lambda(T)) - \omega_{k}^{\delta}(\lambda(T))$$

$$\leq |\omega_{k}(\lambda(T)) - \omega(T)| + |\omega(T) - \omega^{\delta}(T)| + |\omega^{\delta}(T) - \omega_{k}^{\delta}(\lambda(T))|$$

$$\leq d_{T}^{\omega}(\omega, \omega_{k}) + d_{T}^{\delta}(\omega_{k}, \omega^{\delta}) + \omega_{\delta}(T)$$

Using this, together with the claim and Proposition 3.1, we infer that

$$\limsup_{k \to \infty} g_{D_{T}}(\mu_{N}(\omega_{k}), \mu_{N}(\omega_{k}^{\delta})) \leq \omega_{k}(T)e^{(1+\sigma)T+\omega(T)}.$$ 

Now, using Proposition 3.2 together with the claim, we conclude that

$$\limsup_{k \to \infty} g_{D_{T}}(\mu_{N}(\omega_{k}^{\delta}), \mu_{N}(\omega^{\delta})) = 0.$$ 

Hence, letting $k \to \infty$ in (3.10) and using Proposition 3.1, we obtain

$$\limsup_{k \to \infty} g_{D_{T}}(\mu_{N}(\omega_{k}), \mu_{N}(\omega)) \leq 2\omega_{k}(T)e^{(1+\sigma)T+\omega(T)}.$$ 

The previous inequality holds for any $\delta \in A_{\omega}$. It is plain to see that $\inf A_{\omega} = 0$. Hence, letting $\delta \to 0$ with $\delta \in A_{\omega}$ in the previous inequality yields the result.

It remains to prove the claim. Let $\delta \in A_{\omega}$. Since $d_{T}^{\omega}(\omega_{k}, \omega)$ converges to 0 as $k \to \infty$, there is a sequence $(\lambda_{k})_{k \in \mathbb{N}}$ with $\lambda_{k} \in \mathcal{C}_{T}^{\uparrow}$ such that

$$\|\lambda_{k}\|_{T} \to 0 \quad \text{and} \quad \epsilon_{k} := \sum_{u \in [0,T]} |\Delta(\omega_{k} \circ \lambda_{k})(u) - \Delta \omega(u)| \to 0.$$ 

Set $\bar{\omega}_{k} = \omega_{k} \circ \lambda_{k}$. Clearly, $\Delta \bar{\omega}_{k}(u) \leq \epsilon_{k} + \Delta \omega(u)$ and $\Delta \omega(u) \leq \epsilon_{k} + \Delta \bar{\omega}_{k}(u)$, $u \in [0,T]$. Therefore,

$$\omega_{k}^{\delta}(\lambda_{k}(t)) - \omega^{\delta}(t) = \sum_{u \in [0,T] : \Delta \bar{\omega}_{k}(u) \leq \delta} \Delta \bar{\omega}_{k}(u) - \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta} \Delta \omega(u)$$

$$\leq \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta - \epsilon_{k}} \Delta \omega(u) - \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta} \Delta \omega(u)$$

$$= \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta - \epsilon_{k}} (\Delta \bar{\omega}_{k}(u) - \Delta \omega(u)) + \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta - \epsilon_{k}} \Delta \omega(u)$$

$$\leq d_{T}^{\omega}(\omega_{k}, \omega) + \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta - \epsilon_{k}} \Delta \omega(u).$$

Similarly, we obtain

$$\omega^{\delta}(t) - \omega_{k}^{\delta}(\lambda_{k}(t)) = \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta} \Delta \omega(u) - \sum_{u \in [0,T] : \Delta \bar{\omega}_{k}(u) \geq \delta} \Delta \bar{\omega}_{k}(u)$$

$$\leq \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta} \Delta \omega(u) + \sum_{u \in [0,T] : \Delta \bar{\omega}_{k}(u) \geq \delta} (\Delta \omega(u) - \Delta \bar{\omega}_{k}(u))$$

$$\leq \sum_{u \in [0,T] : \Delta \omega(u) \geq \delta + \epsilon_{k}} \Delta \omega(u) + d_{T}^{\omega}(\omega_{k}, \omega).$$

Thus, we deduce that

$$d_{T}^{\omega}(\omega_{k}^{\delta}, \omega^{\delta}) \leq d_{T}^{\omega}(\omega_{k}, \omega) + \sum_{u \in [0,T] : \Delta \omega(u) \in [\delta - \epsilon_{k}, \delta + \epsilon_{k}]} \Delta \omega(u).$$

Since $\delta \in A_{\omega}$, letting $k \to \infty$ in the previous inequality yields $\lim_{k \to \infty} d_{T}^{\omega}(\omega_{k}^{\delta}, \omega^{\delta}) = 0$. Recall that $\omega^{\delta}$ has a finite number of jumps, and hence, the claim follows using Lemma A.1. \[\square\]
4. Wright-Fisher process in random environment: existence, uniqueness and convergence

We start this section with the proof of Theorem 2.2 about the existence and pathwise uniqueness of strong solutions of 4.3.

Proof of Theorem 2.2 (Existence and uniqueness). We prove existence and pathwise uniqueness of strong solutions of 4.3 via 29, Thms. 3.2, 5.1. To this purpose, we first extend 4.3 to an SDE on $\mathbb{R}$ and we write it in the form of 29, Eq. (2.1)]. Note that by Lévy-Itô decomposition, the pure jump subordinator $J$ can be expressed as

$$J(t) = \int_{(0,1)} x N(t, dx), \quad t \geq 0,$$

where $N(ds, dx)$ is a Poisson random measure with intensity measure $\mu$. Hence, defining the functions $a, b : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \times (0, 1) \to \mathbb{R}$ via

$$a(x) := \begin{cases} \sqrt{2x(1-x)}, & \text{for } x \in [0, 1] \\ 0, & \text{otherwise} \end{cases}, \quad g(x, u) := \begin{cases} x(1-x)u, & \text{for } x \in [0, 1] \text{ and } u \in (0, 1), \\ 0, & \text{otherwise} \end{cases},

$$b(x) := \begin{cases} \sigma x(1-x) + \theta \nu_0 (1-x) - \theta \nu_1, & \text{for } x < 0, \\ \theta \nu_0, & \text{for } x > 1, \\ -\theta \nu_1, & \text{for } x > 1, \end{cases}$$

Eq. 4.3 can be extended to the following SDE on $\mathbb{R}$

$$X(t) = X(0) + \int_0^t a(X(s))dB(s) + \int_0^t g(X(s-), u)N(ds, du) + \int_0^t b(X(s))ds, \quad t \geq 0. \quad (4.1)$$

From construction, any solution $(X(t))_{t \geq 0}$ of 4.1 such that $X(t) \in [0, 1]$ for any $t \geq 0$ is a solution of 4.3 and vice-versa.

Note that the functions $a, b, g$ are continuous. Moreover, $b = b_1 - b_2$, where

$$b_1(x) := \theta \nu_0 + \sigma x \quad \text{for } x \in [0, 1], \quad b_1(x) := \theta \nu_0 \quad \text{for } x \leq 0, \quad \text{and} \quad b_1(x) := \theta \nu_0 + \sigma \quad \text{for } x \geq 1,$$

$$b_2(x) := \theta \nu_0 + \sigma x \quad \text{for } x \in [0, 1], \quad b_2(x) := 0 \quad \text{for } x \leq 0, \quad \text{and} \quad b_2(x) := \theta + \sigma \quad \text{for } x \geq 1.$$

In addition, $b_2$ is non-decreasing. Thus, in order to apply 29, Thms. 3.2, 5.1, we only need to verify the sufficient conditions (3.a), (3.b) and (5.a) therein. Condition (3.a) in our case amounts to prove that $x \mapsto b_1(x) + \int_{(0,1)} g(x, u)\mu(du)$ is Lipschitz continuous. In fact, a straightforward calculation shows that

$$|b_1(x) - b_1(y)| + \int_{(0,1)} |g(x, u) - g(y, u)|\mu(du) \leq \left(\sigma + \int_{(0,1)} u\mu(du)\right)|x - y|, \quad x, y \in \mathbb{R},$$

and hence (3.a) follows. Condition (3.b) amounts to prove that $x \mapsto a(x)$ is 1/2-Hölder. We claim that

$$|a(x) - a(y)| \leq 2|x - y|, \quad x, y \in \mathbb{R}. \quad (4.2)$$

One can easily check that 4.2 holds whenever $x \notin (0, 1)$ or $y \notin (0, 1)$. Now, assume that $x, y \in (0, 1)$. If $x + y < 1$, we have

$$|a(x) - a(y)| = \frac{2|x - y|(1 - (x + y))}{\sqrt{2x(1-x) + 2y(1-y)}} \leq \frac{\sqrt{2}|x - y|(1 - (x + y))}{\sqrt{x(1-x) + y(1-y)}} \leq \frac{\sqrt{2}|x - y|(1 - (x + y))}{\sqrt{x + y}(1 - (x + y))} \leq 2|x - y|(1 - (x + y)).$$

Since $a(z) = a(1 - z)$ for all $z \in \mathbb{R}$, the same inequality holds for $x, y \in (0, 1)$ such that $x + y > 1$, and the case $x + y = 1$ is trivial. Hence, the claim is true and condition (3.b) follows. Therefore, 29, Thm. 3.2] yields the pathwise uniqueness for 4.1. Condition (5.a) follows from the fact that the functions $a, b, x \mapsto \int_{(0,1)} g(x, u)^2\mu(du)$ and $x \mapsto \int_{(0,1)} g(x, u)\mu(du)$ are bounded on $\mathbb{R}$. Hence, 29, Thm. 5.1] ensures the existence of a strong solution to 4.1. It remains to show that any solution of 4.1 with $X(0) \in [0, 1]$ is such that $X(t) \in [0, 1]$ for any $t \in [0, 1]$. Sufficient conditions implying such a result are provided in 17.
Prop. 2.1. The conditions on the diffusion and drift coefficients are satisfied, namely, $a$ is 0 outside $[0,1]$ and $b(x)$ is positive for $x \leq 0$ and negative for $x \geq 1$. However, the condition on the jump coefficient, $x_g(x,u) \in [0,1]$ for every $x \in \mathbb{R}$, is not fulfilled. Nevertheless, the proof of [17] Prop. 2.1 works without modifications under the alternative condition $x_g(x,u) \in [0,1]$ for $x \in [0,1]$ and $g(x,u) = 0$ for $x \notin [0,1]$, which is in turn satisfied. This ends the proof.

\[ \square \]

**Lemma 4.1.** The solution of the SDE (1.3) is a Feller process with generator $A$ satisfying for all $f \in C^2([0,1], \mathbb{R})$

\[
Af(x) = x(1-x)f''(x) + (\sigma x(1-x) + \theta \nu_0(1-x) - \theta \nu_1 x)f'(x) + \int_0^1 (f(x + x(1-x)u) - f(x)) \mu(du).
\]

Moreover, $C^\infty([0,1], \mathbb{R})$ is an operator core for $A$.

**Proof.** Since pathwise uniqueness implies weak uniqueness (see [1], Thm. 1), we deduce from [21, Thm. 2.16] that the martingale problem associated to $A$ in $C^\infty([0,1], \mathbb{R})$ is well-posed. Using [32, Prop. 2.2], we infer that $X$ is Feller. The fact that $C^\infty([0,1])$ is a core follows then from [32, Thm. 2.5].

Now, we proceed to prove the annealed convergence of the Moran models towards the Wright–Fisher diffusion stated in Theorem 2.3.

**Proof of Theorem 2.3 (Annealed convergence).** Let $A_N^*$ and $A$ be the infinitesimal generators of the processes $(X^N(Nt))_{t \geq 0}$ and $(X(t))_{t \geq 0}$, respectively. We will prove that, for all $f \in C^\infty([0,1], \mathbb{R})$

\[
\sup_{x \in E_N} |A_N^* f(x) - Af(x)| \to 0 \quad \text{as} \quad N \to \infty.
\]

Provided the claim is true, since $X$ is Feller and $C^\infty([0,1], \mathbb{R})$ is an operator core for $A$ (see Lemma 4.1), the result follows applying [23, Theorem 19.25]. Thus, it remains to prove the claim. To this purpose, we decompose the generator $A$ as $A^1 + A^2 + A^3 + A^4$, where

\[
A^1 f(x) := x(1-x)f''(x), \quad A^2 f(x) := (\sigma x(1-x) + \theta \nu_0(1-x) - \theta \nu_1 x)f'(x),
\]

\[
A^3 f(x) := \int_{0}^{1} (f(x + x(1-x)u) - f(x)) \mu(du), \quad A^4 f(x) := \int_{(\varepsilon_N, 1)} (f(x + x(1-x)u) - f(x)) \mu(du).
\]

Similarly, we write $A_N^* = A_N^1 + A_N^2 + A_N^3 + A_N^4$, where

\[
A_N^1 f(x) := N^2 x(1-x) \left[ f \left( \frac{x + \frac{1}{N}}{N} \right) + f \left( \frac{x - \frac{1}{N}}{N} \right) - 2 f \left( \frac{x}{N} \right) \right],
\]

\[
A_N^2 f(x) := N^2 (\sigma_N x(1-x) + \theta_N \nu_0(1-x)) \left[ f \left( \frac{x + \frac{1}{N}}{N} \right) - f \left( \frac{x}{N} \right) \right] + N^2 \theta_N \nu_1 x \left[ f \left( \frac{x - \frac{1}{N}}{N} \right) - f \left( \frac{x}{N} \right) \right],
\]

\[
A_N^3 f(x) := \int_{0}^{\varepsilon_N} \left( \mathbb{E} [f(x + \varepsilon_N(x,u))] - f(x) \right) \mu(du), \quad A_N^4 f(x) := \int_{(\varepsilon_N, 1)} \left( \mathbb{E} [f(x + \varepsilon_N(x,u))] - f(x) \right) \mu(du),
\]

and $\varepsilon_N(x,u) := \mathcal{H}(N,N(1-x),\beta(Nx,u))/N$, with $\mathcal{H}(N,N(1-x),k) \sim \text{Hyp}(N,N(1-x),k)$ and $\beta(Nx,u) \sim \text{Bin}(Nx,u)$ being independent. The parameter $\varepsilon_N > 0$ will be chosen later in an appropriate way.

Let $f \in C^\infty([0,1], \mathbb{R})$. Note that

\[
\sup_{x \in E_N} |A_N^* f(x) - Af(x)| \leq \sum_{i=1}^{4} \sup_{x \in E_N} |A_i^* f(x) - A_i^1 f(x)|.
\]

Using Taylor expansions of order three around $x$ for $f(x + 1/N)$ and $f(x - 1/N)$, we get

\[
\sup_{x \in E_N} |A_N^1 f(x) - A^1 f(x)| \leq \frac{\|f'''\|_\infty}{3N}.
\]

Similarly, using the triangular inequality and appropriate Taylor expansions of order two yields

\[
\sup_{x \in E_N} |A_N^3 f(x) - A^3 f(x)| \leq \frac{(\sigma_N + \theta_N)\|f'''\|_\infty}{2} + (|\sigma - N\sigma_N| + |\theta - N\theta_N|)\|f''\|_\infty.
\]
For the third term, using that $\mathbb{E}[\xi_N(x,u)] = x(1-x)u$, we get
\[
|A_N^3 f(x)| \leq \|f'\|_\infty \int_{(0,\varepsilon_N)} u \mu(du), \quad x \in [0,1],
\]
and hence,
\[
\sup_{x \in \mathbb{E}_N} |A_N^3 f(x) - A f(x)| \leq \|f'\|_\infty \int_{(0,\varepsilon_N)} u \mu(du).
\]
(4.6)

For the last term, we first note that
\[
|\mathbb{E} [f(x+\xi_N(x,u)) - f(x) - f(x(1-x)u)]| \leq \|f'\|_\infty \mathbb{E} [\xi_N(x,u) - x(1-x)u]
\]
\[
\leq \|f'\|_\infty \mathbb{E} (\xi_N(x,u) - x(1-x)u^2)
\]
\[
\leq \|f'\|_\infty \mathbb{E} (\xi_N(x,u) - x(1-x)u^2) / N.
\]
(4.7)

In the last inequality we used that
\[
\mathbb{E} (\xi_N(x,u) - x(1-x)u^2) = \frac{x(1-x)^2u(1-u)}{N} + \frac{Nx^2(1-x)u^2}{N^2(N+1)} \leq \frac{u}{N},
\]
which is obtained from standard properties of the hypergeometric and binomial distributions. Hence,
\[
\sup_{x \in \mathbb{E}_N} |A_N^3 f(x) - A f(x)| \leq \|f'\|_\infty \int_{(0,\varepsilon_N)} \mathbb{E} (\xi_N(x,u) - x(1-x)u) \mathbb{E} u \mu(du).
\]
(4.8)

Since $\int_{(0,1)} u \mu(du) < \infty$, choosing $\varepsilon_N := 1/\sqrt{N}$, the result follows by plugging (4.4), (4.5), (4.6) and (4.8) in (4.3) and letting $N \to \infty$.

Now, we turn our attention to the asymptotic behavior of a sequence of Moran models in the quenched case. We start with the following lemma that concerns the Moran model with null environment.

**Lemma 4.2.** Let $X^0_N(t) := X_N(0, Nt)$, $t \geq 0$. For any $x \in \mathbb{E}_N$ and $t \geq 0$, we have
\[
\mathbb{E}_x [(X^0_N(t) - x)^2] \leq \left( \frac{1}{2} + N(\sigma_N + 3\theta_N) \right) t,
\]
and
\[
-N\theta_N \nu_1 t \leq \mathbb{E}_x [X^0_N(t) - x] \leq N(\sigma_N + \nu_1) t.
\]

**Proof.** Fix $x \in \mathbb{E}_N$ and consider the functions $f_x, g_x : \mathbb{E}_N \to [0,1]$ defined via $f_x(z) := (z-x)^2$ and $g_x(z) := z-x, z \in \mathbb{E}_N$. The process $X^0_N$ is a Markov chain with generator $A^0_N := A_N + A_N^0$, where $A_N$ and $A_N^0$ are defined in the proof of Theorem 2.3. Moreover, for every $z \in \mathbb{E}_N$, we have
\[
A_N^0 f_x(z) = 2z(1-z) + N \left[ (\sigma_N z + \theta_N \nu_0)(1-z) \left( 2(z-x) + \frac{1}{N} \right) + N\theta_N \nu_1 z \left( 2(x-z) + \frac{1}{N} \right) \right]
\]
\[
\leq \frac{1}{2} + N(\sigma_N + 3\theta_N),
\]
and
\[
A_N^0 g_x(z) = N \{ (\sigma_N z + \theta_N \nu_0)(1-z) - \theta_N \nu_1 z \} \in [-N\theta_N \nu_1, N(\sigma_N + \theta_N \nu_0)].
\]
Hence, Dynkin’s formula applied to $X^0_N$ with the function $z \in \mathbb{E}_N \mapsto (z-x)^2$ leads to
\[
\mathbb{E}_x [(X^0_N(t) - x)^2] = \int_0^t \mathbb{E}_s [A^0_N f_x(X^0_N(s))] ds \leq \left( \frac{1}{2} + N(\sigma_N + 3\theta_N) \right) t.
\]
Similarly, applying Dynkin’s formula to $X^0_N$ with $g_x$, we obtain
\[
\mathbb{E}_x [X^0_N(t) - x] = \int_0^t \mathbb{E}_s [A^0_N g_x(X^0_N(s))] ds \in [-N\theta_N \nu_1 t, N(\sigma_N + \theta_N \nu_0) t],
\]
which ends the proof.
Now, note that using the two previous inequalities and the tower property of the conditional expectation, we get

\[ \mathbb{E}[(X_N^n(t) - X_N^n(s)) | \mathcal{F}_s^N] \leq c \sum_{u \in [s,t]} \Delta \omega(u) + C(t-s), \]

where \( (\mathcal{F}_s^N)_{s \geq 0} \) denotes the natural filtration associated to the process \( X_N^n \). If the claim is true, then condition A2) is satisfied with \( \beta = 2 \) and \( F_N(t) = F(t) = c \sum_{u \in (0,t]} \Delta \omega(u) + Ct \), and the result follows from Lemma 4.2 therein. The rest of the proof is devoted to prove the claim.

For \( x \in E_N \) and \( t \geq 0 \), we set

\[ \psi_x(\omega, t) := \mathbb{E}_x[(X_N^n(t) - x)^2]. \]

From the definition of \( X_N^n \), it follows that, for any \( 0 \leq s < t \)

\[ \mathbb{E}[(X_N^n(t) - X_N^n(s))^2 | \mathcal{F}_s^N] = \psi_{X_N^n(s)}(\omega, t-s) \tag{4.9} \]

where \( \omega_s(\cdot) := \omega(s + \cdot) \).

Let \( 0 \leq s < t \). We split the proof of the claim in three cases.

**Case 1:** \( \omega \) has no jumps in \( (s,t] \). In particular, \( \omega_s \) has no jumps in \( (0,t-s] \). Hence, restricted to \( (0,t-s] \), \( X_N^n \) has the same distribution as \( X_N^0 \). Using Lemma 4.1 with \( x = X_N^n(s) \) and plugging the result in (4.9), we infer that in this case, the claim holds for any \( c \geq 1 \) and \( C \geq c_1 := 1/2 + \sup_{N \in \mathbb{N}} (N(\sigma_N + \rho \theta_N)). \)

**Case 2:** \( \omega \) has \( n \) jumps in \( (s,t] \). We denote by \( t_1, \ldots, t_n \in (s,t] \) the jumping times of \( \omega \) in \( (s,t] \) in increasing order. We set \( t_0 := s \) and \( t_{n+1} = t \). For any \( i \in [n+1] \) and any \( r \in (t_{i-1}, t_i) \), \( \omega \) has no jumps in \( (t_{i-1}, r] \). In particular, \( t_i \) falls into case 1. Using the claim and letting \( r \to t_i \), we obtain

\[ \mathbb{E}[(X_N^n(t_i) - X_N^n(t_{i-1}))^2 | \mathcal{F}_{t_{i-1}}^N] \leq c(t_i - t_{i-1}). \]

Moreover,

\[ \mathbb{E}[(X_N^n(t_i) - X_N^n(t_{i-1}))^2 | \mathcal{F}_{t_{i-1}}^N] \leq \mathbb{E} \left( \frac{B_N \Delta \omega(t_i)}{N} \right)^2 \leq \Delta \omega(t_i). \]

Using the two previous inequalities and the tower property of the conditional expectation, we get

\[ \mathbb{E}[(X_N^n(t_i) - X_N^n(t_{i-1}))^2 | \mathcal{F}_{t_{i-1}}^N] \leq 2C_1(t_i - t_{i-1}) + 2\Delta \omega(t_i). \tag{4.10} \]

Now, note that

\[ (X_N^n(t) - X_N^n(s))^2 = \left( \sum_{i=0}^{n} (X_N^n(t_{i+1}) - X_N^n(t_i)) \right)^2 \]

\[ = \sum_{i=0}^{n} (X_N^n(t_{i+1}) - X_N^n(t_i))^2 + 2 \sum_{i=0}^{n} (X_N^n(t_{i+1}) - X_N^n(t_i)) (X_N^n(t_i) - X_N^n(s)). \]

Using Eq. (4.10), we see that

\[ \mathbb{E} \left[ \sum_{i=0}^{n} (X_N^n(t_{i+1}) - X_N^n(t_i))^2 | \mathcal{F}_s^N \right] \leq 2C_1(t-s) + 2 \sum_{i=1}^{n} \Delta \omega(t_i). \]

Moreover, we have

\[ \mathbb{E}[(X_N^n(t_{i+1}) - X_N^n(t_i))(X_N^n(t_i) - X_N^n(s)) | \mathcal{F}_{t_i}^N] = \varphi_{X_N^n(s), X_N^n(t_i)}(\omega_{t_i}, t_{i+1} - t_i). \]

where \( \varphi_{x,y}(\omega, t) := (y-x)\mathbb{E}_x[X_N^n(t) - y] \). Since, for any \( r \in (t_i, t_{i+1}) \), \( \omega_{t_i} \) has no jumps in \( (t_i, r] \), we can apply Lemma 4.2 to obtain

\[ \varphi_{x,y}(\omega_{t_i}, r-t_i) \leq N((\sigma_N + \theta_N \nu_0) \lor \theta_N \nu_1)(r-t_i) \]
Letting $r \to t_{i+1}$ and noting that $(y - x)E_y[X^\omega_N((t_{i+1} - t_i) - X^\omega_N((t_{i+1} - t_i) - x)] \leq \Delta^\omega(t_{i+1})$, we conclude that
\[ \varphi_x(y, t_{i+1} - t_i) \leq N((\sigma_N + \theta_N\nu_0) + \theta_N\nu_1)(t_{i+1} - t_i) + \Delta^\omega(t_{i+1}). \]
All together, we obtain
\[ E[(X^\omega_N(t) - X^\omega_N(s))^2 | F^N_s] \leq C_2(t - s) + 3 \sum_{i=1}^{n} \Delta^\omega(t_i), \]
where $C_2 := 2C_1 + \sup_{N \in \mathbb{N}} N((\sigma_N + \theta_N\nu_0) + \theta_N\nu_1)$. Hence, the claim holds for any $C \geq C_2$ and $c \geq 3$.

**Case 3:** $\omega$ has infinitely many jumps in $(s, t]$. For any $\delta$, we consider $\omega^\delta$ as in Section 3 and we couple the processes $X^\omega_N$ and $X^\omega_N$ as in the proof of Proposition 3.1. Note that $\omega^\delta$ has only a finite number of jumps in any compact interval, thus $\omega^\delta$ falls into case 2. Moreover, we have
\[ \psi_x(\omega, t) \leq 2E_x[(X^\omega_N(t) - X^\omega_N(t))^2] + 2E_x[(X^\omega(t) - x)^2] \]
\[ \leq 2e^{N\sigma(t)} \sum_{u \in [0, t]: \Delta^\omega(u) < \delta} \Delta^\omega(u) + 2E_x[(X^\omega(t) - x)^2], \]
where in the last inequality we use Proposition 3.1. Now, using the claim for $X^\omega_N$ and the previous inequality, we obtain
\[ E[(X^\omega_N(t) - X^\omega_N(s))^2 | F^N_s] \leq e^{N\sigma(t-s) + \omega(t-s)} \sum_{u \in [s, t]: \Delta^\omega(u) < \delta} \Delta^\omega(u) + 2E_x[(X^\omega(t) - x)^2], \]
We let $\delta \to 0$ and conclude that the claim holds for any $C \geq 2C_2$ and $c \geq 6$.

Now, we proceed to prove the quenched convergence of the sequence of Moran models to the Wright–Fisher diffusion, under the assumption that the environment is simple.

**Proof of Theorem 2.4 (Quenched convergence).** Let $B := (B(t))_{t \geq 0}$ be a standard Brownian motion. We denote by $X^0$ the unique strong solution of (1.3) with null environment associated to $B$.

For any simple environment $\omega$, we set $X^\omega_N(t) := X^\omega_N(\omega; Nt)$ for $t \geq 0$ and $N \geq 1$. Note that Theorem 2.3 implies in particular that $X^\omega_N$ converges to $X^0$ as $N \to \infty$.

Now, assume that $\omega$ is simple, but not constant equal to 0. We denote by $T_\omega$ the set of jumping times of $\omega$ in $(0, \infty)$. For any $0 < i < |T_\omega| + 1$, we denote by $t_i := t_i(\omega) \in T_\omega$ the time of the $i$-th jump of $\omega$. We set $t_0 = 0$ and $t_{|T_\omega|+1} = \infty$. Therefore, we need to prove that, under the assumptions of the Theorem, we have
\[ (X^\omega_N(t))_{t \geq 0} \xrightarrow{\text{d}}_{N \to \infty} (X^\omega(t))_{t \geq 0}, \]
where the process $X^\omega$ is defined as follows.

- $X^\omega(0) = X^\omega_0$.
- For $i \in \mathbb{N}$ with $i \leq |T_\omega| + 1$, the restriction of $X^\omega$ to the interval $(t_{i-1}, t_i)$ is given by a version of $X^0$ started at $X^\omega(t_{i-1})$.
- For $0 < i < |T_\omega| + 1$, conditionally on $X^\omega(t_{i-1})$,
\[ X^\omega(t_i) = X^\omega(t_{i-1}) + X^\omega(t_{i-1})(1 - X^\omega(t_{i-1})) \Delta^\omega(t_i). \]

Since the sequence $(X^\omega_N)_{N \in \mathbb{N}}$ is tight, it is enough to prove the convergence at the level of the finite dimensional distributions. More precisely, we will prove by induction on $i \in \mathbb{N}$ with $i \leq |T_\omega| + 1$ that for any finite set $I \subset [0, t_i)$, we have
\[ ((X^\omega_N(t))_{t \in I}, X^\omega_N(t_i)) \xrightarrow{\text{d}}_{N \to \infty} ((X^\omega(t))_{t \in I}, X^\omega(t_i)). \]

The result for $i = 1$ follows from Theorem 2.3 and the fact that $X^\omega(t_1) = X_N^\omega(t_1)$ and $X^\omega(t_1) = X^0(t_1)$ almost surely. Now, assume that the result is true for some $i < |T_\omega| + 1$ and let $I \subset (0, t_{i+1})$. Without loss of generality we assume that $I = \{s_1, \ldots, s_k, t_i, r_1, \ldots, r_m\}$, with $s_1 < \cdots < s_k < t_i < r_1 < \cdots < r_m$. We also assume that $i < |T_\omega|$, the other case, i.e. $i = |T_\omega| < \infty$, follows using an analogous argument.
Let $F : [0,1]^{k+1} \to \mathbb{R}$ be a Lipschitz function with $\|F\|_{BL} \leq 1$. Note that

$$
\mathbb{E} \left[ F((X_N^x(s_j))_{j=1}^k, X_N^x(t)) \right] = \mathbb{E} \left[ F((X_N^x(s_j))_{j=1}^k, X_N^x(t_i) + \xi_N(X_N^x(t_i), \Delta \omega(t_i))) \right],
$$

where for $x \in E_N$ and $u \in (0,1)$, $\xi_N(x,u) := \mathcal{H}(N,N(1 - x), \beta(Nx,u))/N$ with $\mathcal{H}(N,N(1 - x), k) \sim \text{Hyp}(N,N(1 - x), k)$, $k \in [N]_0$, and $\beta(Nx,u) \sim \text{Bin}(Nx,u)$ being independent between them and independent of $X_N^x$. Now, set

$$
D_N := \mathbb{E} \left[ F((X_N^x(s_j))_{j=1}^k, X_N^x(t_i) + \xi_N(X_N^x(t_i), \Delta \omega(t_i))) \right] - \mathbb{E} \left[ F((X_N^x(s_j))_{j=1}^k, X_N^x(t_i) + \xi_N(X_N^x(t_i), \Delta \omega(t_i))) \right].
$$

Using that $\|F\|_{BL} \leq 1$ and (4.7), we see that $|D_N| \leq \sqrt{\Delta \omega(t_i)}/N \to 0$ as $N \to \infty$. Therefore, the induction hypothesis yields

$$
\mathbb{E} \left[ F((X_N^x(s_j))_{j=1}^k, X_N^x(t_i)) \right] = D_N + \mathbb{E} \left[ F((X_N^x(s_j))_{j=1}^k, X_N^x(t_i) + \xi_N(X_N^x(t_i), \Delta \omega(t_i))) \right] \to \mathbb{E} \left[ F((X^x(s_j))_{j=1}^k, X^x(t_i)) \right].
$$

Therefore,

$$(X^x(s_j))_{j=1}^k, X^x(t_i)) \to_{N \to \infty} (X^x(s_j))_{j=1}^k, X^x(t_i). \quad (4.11)$$

Let $G : [0,1]^{k+m+2} \to \mathbb{R}$ be a Lipschitz function with $\|G\|_{BL} \leq 1$. For $x \in E_N$, define

$$
H_N(z,x) := E_x[G(z,x,(X_N^x(r_j - t_i))_{j=1}^m, X_N^x(t_{i+1} - t_i))], \forall z \in \mathbb{R}^k.
$$

Note that

$$
E[G((X_N^x(s_j))_{j=1}^k, X_N^x(t_i), (X_N^x(r_j))_{j=1}^m, X_N^x(t_{i+1} - t_i))] = E[H_N((X_N^x(s_j))_{j=1}^k, X_N^x(t_i))]. \quad (4.12)
$$

Similarly, for $x \in [0,1]$, define

$$
H(z,x) := E_z[G(z,x,(X^0(r_j - t_i))_{j=1}^m, X^0(t_{i+1} - t_i))], \forall z \in \mathbb{R}^k,
$$

and note that

$$
E[G((X^x(s_j))_{j=1}^k, X^x(t_i), (X^x(r_j))_{j=1}^m, X^x(t_{i+1} - t_i))] = E[H((X^x(s_j))_{j=1}^k, X^x(t_i))]. \quad (4.13)
$$

Using (4.11) and the Skorokhod representation theorem, we can assume that $((X_N^x(s_j))_{j=1}^k, X_N^x(t_i))_{N \geq 1}$ and $((X^x(s_j))_{j=1}^k, X^x(t_i))$ live in the same probability space and that the convergence holds almost surely. In particular, we can write

$$
|E[H_N((X_N^x(s_j))_{j=1}^k, X_N^x(t_i))]| - E[H((X^x(s_j))_{j=1}^k, X^x(t_i))] | \leq R^1_N + R^2_N,
$$

where

$$
R^1_N := |E[H_N((X_N^x(s_j))_{j=1}^k, X_N^x(t_i))]| - E[H_N((X_N^x(s_j))_{j=1}^k, X_N^x(t_i))]|,
$$

$$
R^2_N := |E[H_N((X_N^x(s_j))_{j=1}^k, X_N^x(t_i))]| - E[H((X^x(s_j))_{j=1}^k, X^x(t_i))]|.
$$

Using that $\|G\|_{BL} \leq 1$, we obtain

$$
R^1_N \leq \sum_{j=1}^k E[|X_N^x(s_j) - X^x(s_j)|] \to 0. \quad (4.15)
$$

Moreover, since $X_N^x(t_i)$ converges to $X^x(t_i)$ almost surely, we conclude using Theorem 2.3 that, for any $z \in [0,1]^k$, $H_N(z,X_N^x(t_i))$ converges to $H(z,X^x(t_i))$ almost surely. Therefore, using dominated convergence theorem, we conclude that

$$
R^2_N \to 0. \quad (4.16)
$$

Plugging (4.15) and (4.16) in (4.14) and using (4.12) and (4.13) completes the proof. \qed
5. Type distribution and ancestral type distribution: the annealed case

5.1. Type distribution: the killed ASG. In this section, we prove Theorem 2.9 about the moment duality between the diffusion $X$ and the killed ASG, and Theorem 2.10 about the moments of the stationary distribution of $X$.

Proof. of Theorem 2.9 Let $H : [0, 1] \times \mathbb{N} \cup \{0, \} \text{ defined via } H(x, n) = (1 - x)^n$. By [22, Prop. 1.2], we only need to show that $G_X H(\cdot, n)(x) = G_R H(x, \cdot)(n)$ for all $n \in \mathbb{N}$, where $G_R$ denotes the generator of the process $R$.

Note first that

$G_X H(\cdot, n)(x) = n(n - 1)x(1 - x)^{n-1} - (\sigma x(1 - x) + \theta \nu_0(1 - x) - \theta \nu_1 x) n(1 - x)^{n-1}$

$+ (1 - x)^n \int ((1 - xz)^n - 1)n(1 - x)^{-1} \mu(dz)$. \hfill (5.1)

In addition,

$G_R H(x, \cdot)(n) = (n(n - 1) + n \theta \nu_1)((1 - x)^{n-1} - (1 - x)^n) - n \theta \nu_0(1 - x)^n + n \sigma(1 - x)^{n+1} - (1 - x)^n$

$+ \sum_{k=1}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) \int_{(0,1)} y^k(1 - y)^{n-k} \mu(dy)((1 - x)^{n+k} - (1 - x)^n)$

$= n(n - 1)x(1 - x)^{n-1} - (\sigma x(1 - x) + \theta \nu_0(1 - x) - \theta \nu_1 x) n(1 - x)^{n-1}$

$+ (1 - x)^n \sum_{k=1}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) \int_{(0,1)} y^k(1 - y)^{n-k} \mu(dy)((1 - x)^k - 1)$. \hfill (5.2)

Moreover, we have

$\sum_{k=1}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) \int_{(0,1)} y^k(1 - y)^{n-k} \mu(dy)((1 - x)^k - 1) = \int_{(0,1)} \sum_{k=1}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) (((1 - x)y)^k - y^k)(1 - y)^{n-k} \mu(dy)$

$= \int_{(0,1)} ((1 - x)^n) \mu(dy)$. \hfill (5.3)

The desired result follows by plugging (5.3) in (5.2) and comparing the result with (5.1).

Proof of Theorem 2.10. We first show the convergence in distribution of $X(T)$ as $T \to \infty$ towards a limit law on $[0, 1]$. Since $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$, Theorem 2.9 implies that, for any $x \in [0, 1]$ the limit of $E[(1 - X(T))^n | X(0) = x]$ as $T \to \infty$ exists and satisfy

$\lim_{T \to \infty} E[(1 - X(T))^n | X(0) = x] = w_n, \quad n \in \mathbb{N}_0.$ \hfill (5.4)

Recall that on $[0, 1]$ probability measures are completely determined by their moments and convergence of positive entire moments implies convergence in distribution. Therefore, Eq. (5.4) implies that there is a probability distribution $\pi_X$ on $[0, 1]$, such that, for any $x \in [0, 1]$, conditionally on $\{X(0) = x\}$, the law of $X(T)$ converges in distribution to $\pi_X$ as $T \to \infty$ and

$w_n = \int_{[0,1]} (1 - z)^n \pi_X(dz), \quad n \in \mathbb{N}.$

Using dominated convergence, the convergence of the law of $X(T)$ towards $\pi_X$ as $T \to \infty$ extends to any initial distribution. As a consequence of this and the Markov property of $X$, it follows that $X$ admits a unique stationary distribution, which is given by $\pi_X$.

It only remains to prove (2.6). For this, we do a first step decomposition for the probability of absorption in 0 of the process $R$, and obtain

$\ell_n w_n = n \sigma w_{n+1} + n(\theta \nu_1 + n - 1) w_{n-1} + \sum_{k=1}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) \sigma_{n,k} w_{n+1},$
where \( t_n := n(\sigma + \theta + n - 1) + \sum_{k=1}^{n} \binom{n}{k}\sigma_{n,k} \). The result follows dividing both sides of the previous by \( n \) and rearranging terms.

\[ \]

5.2. Ancestral type distribution: the pruned lookdown ASG. This section is devoted to prove the results stated in Section 2.6.2 about the ancestral type distribution in the annealed case.

Proof of Lemma 2.17 (Positive recurrence). Since the Markov chain is irreducible, it is enough to prove that the state 1 is positive recurrent for \( L \). The latter is clearly true if \( \theta_0 > 0 \), because in this case the hitting time of one is upper bounded by an exponential random variable with parameter \( \theta_0 \). Now assume that \( \theta = 0 \) (the case \( \theta_0 = 0 \) and \( \theta_1 > 0 \), can be easily reduced to this case). We proceed in a similar way as in [16, Proof of Lem. 2.3]. Consider the function \( f : \mathbb{N} \to \mathbb{R}_+ \) defined via

\[
\begin{align*}
\quad f(n) & := n^{-1} \sum_{i=1}^{n-1} \frac{1}{i} \ln \left( 1 + \frac{1}{i} \right).
\end{align*}
\]

Note that the function \( f \) is bounded. Note also that, for \( n > 1 \)

\[
\begin{align*}
n(n-1)(f(n-1) - f(n)) &= -n \ln \left( 1 + \frac{1}{n-1} \right) \leq -1.
\end{align*}
\]

The previous inequality uses the fact that \( \ln(1-h) \leq -h \) for \( h \in (0,1) \). For any \( \varepsilon > 0 \), set \( m_0(\varepsilon) = \lfloor 1/\varepsilon \rfloor + 1 \). Note that for \( n > m_0(\varepsilon) \)

\[
\begin{align*}
n(n+1) - f(n) &= n \sum_{i=n}^{n+1} \frac{1}{i} \ln \left( 1 + \frac{1}{i} \right) \leq n \ln \left( 1 + \frac{1}{n} \right) \leq \varepsilon.
\end{align*}
\]

Hence, for \( n \geq m_0(\varepsilon) \),

\[
\begin{align*}
G_L f(n) & \leq -1 + \frac{\varepsilon}{n} \sum_{i=1}^{n} \binom{n}{i}\sigma_{n,i} + \varepsilon = -1 + \varepsilon \left( \int_{(0,1)} y \mu(dy) + \sigma \right).
\end{align*}
\]

Now, set \( m_0 := m_0(\varepsilon_*) \), where \( \varepsilon_* := \left( 2 \int_{(0,1)} y \mu(dy) + 2\sigma \right)^{-1} \). Thus, for any \( n \geq m_0 \) we have

\[
\begin{align*}
G_L f(n) & \leq -1/2.
\end{align*}
\]

Define \( T_{m_0} := \inf\{ \beta > 0 : L(\beta) < m_0 \} \). Applying Dynkin’s formula to \( L \) with the function \( f \) and the stopping time \( T_{m_0} \land k, k \in \mathbb{N} \), we obtain

\[
\begin{align*}
\mathbb{E}[f(L(T_{m_0} \land k)) | L(0) = n] &= f(n) + \mathbb{E} \left[ \int_{0}^{T_{m_0} \land k} G_L f(L(\beta)) d\beta | L(0) = n \right]
\end{align*}
\]

Therefore, for \( n \geq m_0 \)

\[
\begin{align*}
\mathbb{E}[f(L(T_{m_0} \land k)) | L(0) = n] & \leq f(n) - \frac{1}{2} \mathbb{E}[T_{m_0} \land k | L(0) = n].
\end{align*}
\]

Hence, using that \( f \) is non-negative, we get

\[
\begin{align*}
\mathbb{E}[T_{m_0} \land k | L(0) = n] & \leq 2f(n).
\end{align*}
\]

Therefore, letting \( k \to \infty \) in the previous inequality yields, for all \( n \geq m_0 \),

\[
\begin{align*}
\mathbb{E}[T_{m_0} | L(0) = n] & \leq 2f(n) < \infty.
\end{align*}
\]

Since \( L \) is irreducible, it follows by standard arguments that \( L \) is positive recurrent.

\[ \]

Proof of Proposition 2.17. The first part of the statement is a straightforward consequence of Lemma 2.17 and the fact that we assign types to the \( L(T) \) lines present in the pruned LD-ASG at time \( T \) according to independent Bernoulli random variables with parameter \( x \). Since \( L \) is positive recurrent, we have convergence in distribution of the law of \( L(T) \) towards the stationary distribution \( \pi_L \). In particular, we infer from Eq. (2.19) that the limit \( h(x) \) of \( h_T(x) \) as \( T \to \infty \) exists and satisfies

\[
\begin{align*}
h(x) = 1 - \mathbb{E}[(1-x)^{L_N}] = 1 - \sum_{\ell=1}^{\infty} \mathbb{P}(L_\ell = 0)(1-x)^{\ell} = \sum_{\ell=0}^{\infty} \mathbb{P}(L_\ell > \ell)(1-x)^{\ell} - \sum_{\ell=1}^{\infty} \mathbb{P}(L_\ell > \ell-1)(1-x)^{\ell-1},
\end{align*}
\]

\[ \]
and Eq. 2.20 follows.

**Proof of Corollary 2.18** Since \( \theta = 0 \), the block-counting processes \( R \) and \( L \) are equal. Hence, combining Proposition 2.17 and Theorem 2.4 applied to \( n = 1 \), we obtain

\[
h_T(x) = \mathbb{E}[X(T) \mid X(0) = x],
\]

which proves the first part of the statement. Moreover, for \( \theta = 0 \), \( X \) is a bounded super-martingale, and hence \( X(T) \) has almost surely a limit as \( T \to \infty \), which we denote by \( X(\infty) \). Letting \( T \to \infty \), in the identity (5.3) yields

\[
h(x) = \mathbb{E}[X(\infty) \mid X(0) = x].
\]

Moreover, using Theorem 2.4 with \( n = 2 \), we get

\[
\mathbb{E}[(1 - X(T))^2 \mid X(0) = x] = \mathbb{E}[(1 - x)^{L(T)} \mid L(0) = 2].
\]

Letting \( T \to \infty \) and using that \( L \) is positive recurrent, we obtain

\[
\mathbb{E}[(1 - X(\infty))^2 \mid X(0) = x] = 1 - h(x).
\]

Plugging (5.6) in the previous identity yields the desired result.

The proof of Theorem 2.19 providing the characterization of the tail probabilities \( \mathbb{P}(L_\infty > n) \) via a linear system of equation, is based on the notion of Siegmund duality. Let us consider the process \( D := (D(\beta))_{\beta \geq 0} \) with rates

\[
q_D(i, j) := \begin{cases} 
(i - 1)(\sigma + \sigma_{i-1,1}) & \text{if } j = i - 1, \\
(i - 1)\theta \nu_1 + \iota(i - 1) & \text{if } j = i + 1, \\
\gamma_{i,j} - \gamma_{i,j-1} & \text{if } 1 \leq j < i - 1, \\
(i - 1)\theta \nu_0 & \text{if } j = \dagger,
\end{cases}
\]

where \( \gamma_{i,j} := \sum_{k=i-j}^{j} \left(\begin{smallmatrix} i \\ k \end{smallmatrix}\right) \sigma_{j,k} \) and \( \dagger \) is a cemetery point. Note that 1 and \( \dagger \) are absorbing states of \( D \).

**Lemma 5.1** (Siegmund duality). The processes \( L \) and \( D \) are Siegmund dual, i.e. for all \( \ell, d \in \mathbb{N} \) and \( t \geq 0 \), we have

\[
\mathbb{P}(L(\beta) \geq d \mid L(0) = \ell) = \mathbb{P}(\ell \geq D(\beta) \mid D(0) = d).
\]

**Proof.** We consider the function \( H : \mathbb{N} \times \mathbb{N} \cup \{\dagger\} \to \{0, 1\} \) defined via \( H(\ell, d) := 1_{\ell \geq d} \) and \( H(\ell, \dagger) := 0 \), \( \ell, d \in \mathbb{N} \). Let \( G_L \) and \( G_D \) be the infinitesimal generators of \( L \) and \( D \), respectively. By [22, Prop. 1.2] we only have to show that \( G_L H(\cdot, d)(\ell) = G_D H(\ell, \cdot)(d) \) for all \( \ell, d \in \mathbb{N} \). From construction, we have

\[
G_L H(\cdot, d)(\ell) = \sigma \ell 1_{\ell+1=d} - (\ell - 1)(\ell + \theta \nu_1) 1_{\ell=d} - \theta \nu_0 \sum_{j=1}^{\ell-1} 1_{j \leq d \leq \ell} + \sum_{k=1}^{\ell} \left(\begin{smallmatrix} \ell \\ k \end{smallmatrix}\right) \sigma_{\ell,k} 1_{\ell \leq d \leq \ell+k} = \gamma \ell 1_{\ell+1=d} - (\ell - 1)(\ell + \theta \nu_1) 1_{\ell=d} - \theta \nu_0 (d-1) 1_{d \leq \ell} + \gamma_\ell 1_{\ell < d}.
\]

Similarly, we have

\[
G_D H(\ell, \cdot)(d) = \sigma(d-1)\ell 1_{d-1=\ell} - (d-1)(d + \theta \nu_1) 1_{\ell=d} - \theta \nu_0 (d-1) 1_{d \leq \ell}
\]

\[
+ \sum_{j=1}^{d-1} (\gamma_{d,j} - \gamma_{d,j-1}) 1_{j \leq \ell < d}.
\]

In addition, using summation by parts yields

\[
\sum_{j=1}^{d-1} (\gamma_{d,j} - \gamma_{d,j-1}) 1_{j \leq \ell < d} = \gamma_{d,d-1} 1_{d-1=\ell} + \gamma_\ell 1_{\ell \leq d-2} = \gamma_\ell 1_{\ell < d}.
\]

The result follows plugging (5.9) in (5.8) and comparing with (5.7).

Now, we can prove Theorem 2.19.
Proof of Theorem 2.11. From Lemma 5.1 we infer that $a_n = d_{n+1}$, where
\[ d_n := \mathbb{P}(\exists \beta > 0 : D(\beta) = 1 \mid D(0) = n), \quad n \geq 1. \]
Applying a first step decomposition to the process $D$, we obtain
\[ T_n d_n = (n-1)\sigma d_{n-1} + (n-1)(\theta v_1 + n)d_{n+1} + \sum_{j=1}^{n-1} (\gamma_{n,j} - \gamma_{n,j-1}) d_j, \quad n > 1, \tag{5.10} \]
where $T_n := (n-1)[\sigma + \theta + n] + \sum_{k=1}^{n-1} \binom{n-1}{k} \sigma d_{n-1,k}$. Using summation by parts and rearranging terms in (5.10) yields
\[ (\sigma + \theta + n)d_n = \sigma d_{n-1} + (\theta v_1 + n)d_{n+1} + \frac{1}{n-1} \sum_{j=1}^{n-1} \gamma_{n,j}(d_j - d_{j+1}), \quad n > 1, \tag{5.11} \]
The result follows.

6. Type distribution and ancestral type distribution: the quenched case

In this section we assume that the environment $\omega$ is fixed and simple.

6.1. Type distribution: the killed ASG. This section is devoted to the proofs of the results presented in Section 5.5.2. In what follows we assume that $\omega$ is a simple environment. Two ingredients for the proof of Theorem 2.11 are 1) a moment duality between the jumping times and 2) a moment duality at the jumping times, these are the object of the two following lemmas.

Lemma 6.1 (Quenched moment duality between the jumps). Let $0 \leq s < t < T$ and assume that $\omega$ has no jumps in $(s, t)$. For all $x \in [0, 1]$ and $n \in \mathbb{N}$, we have
\[ \mathbb{E}^\omega\left[(1 - X(\omega, t -))_n \mid X(\omega, s) = x\right] = \mathbb{E}^\omega\left[(1 - x)^{R(t, \omega)} \mid R(t, \omega, 0) = n\right]. \]
Recall that if $\omega$ has no jump at $t$, then $X(\omega, t-) = X(\omega, t)$ and $R(t, \omega, 0) = R(t, \omega, 0)$.

Proof. In $(s, t)$, the dynamic of $X(\omega, \cdot)$ and $R(t, \omega, \cdot)$ are the same as in the annealed case with $\mu = 0$. Therefore the result follows from Theorem 2.9 applied with $\mu = 0$.

Lemma 6.2 (Quenched moment duality at jumps). For all $x \in [0, 1]$ and $n \in \mathbb{N}$, if $\omega$ has a jump at time $t < T$, then we have
\[ \mathbb{E}^\omega\left[(1 - X(\omega, t))_n \mid X(\omega, t-) = x\right] = \mathbb{E}^\omega\left[(1 - x)^{R(t, \omega)} \mid R(t, \omega, T -) = n\right]. \]

Proof. On the one hand, the equation satisfied by $X(\omega, \cdot)$, that is, \ref{eq:quenched-moment-duality-jumps}, tells us that we have almost surely $X(\omega, t) = X(\omega, t-) + X(\omega, t-) (1 - X(\omega, t-)) \Delta \omega(t)$ so that
\[ \mathbb{E}^\omega\left[(1 - X(\omega, t))_n \mid X(\omega, t-) = x\right] = [1 - x (1 + (1 - x) \Delta \omega(t))]^n = \left[1 - x - \Delta \omega(t)x + \Delta \omega(t)x^2\right]^n. \tag{6.1} \]
One the other hand, recall that, conditionally on $\{R(t, \omega, T -) = n\}$, we have $R(t, \omega, T -) \sim n + Y$ where $Y \sim \text{Bin}(n, \Delta \omega(t))$. Therefore
\[ \mathbb{E}^\omega\left[(1 - x)^{R(t, \omega)} \mid R(t, \omega, T -) = n\right] = \mathbb{E}^\omega\left[(1 - x)^{n + Y}\right] = (1 - x)^n \left[(1 - \Delta \omega(t)) + \Delta \omega(t)(1 - x)\right]^n = \left[1 - x - \Delta \omega(t)x + \Delta \omega(t)x^2\right]^n. \tag{6.2} \]
The combination of \ref{eq:quenched-moment-duality-jumps} and \ref{eq:quenched-moment-duality-at-jumps} yields the result.

Proof of Theorem 2.11 (quenched moment duality). Let $\omega$ be a simple environment. Let $\{T_i\}_{i=1}^\infty$ be the increasing sequence of jumping times of $\omega$ in $[0, T]$. We assume that $\omega$ does not jump at time 0 nor at time $T$ (if $\omega$ does jump at one or both of these times, the following proof can be easily adapted). We set $T_0 := 0$ for convenience. Let $(X(\omega, s))_{s \in [0, T]}$ and $(R(t, \omega, \beta))_{\beta \in [0, T]}$ be independent realizations of the diffusion and the killed ASG, respectively. Since $\omega$ does not jump at $T$, this implies in particular that

$X(\omega, T) = X(\omega, T^-)$ and that $R_T(\omega, 0-) = R_T(\omega, 0)$. Hence, partitioning on the values of $X(\omega, T_N)$ and using Lemma 6.2 we get

$$E^\omega [(1 - X(\omega, T))^n | X(\omega, 0) = x] = E^\omega [(1 - X(\omega, T^-))^n | X(\omega, 0) = x]$$

$$= \int_0^1 E^\omega[(1 - X(\omega, T^-))^n | X(\omega, T_N) = y] \times P^\omega(X(\omega, T_N) \in dy | X(\omega, 0) = x)$$

$$= \int_0^1 E^\omega[(1 - y)^{R_T(\omega,(T^-)-)} | R_T(\omega, 0) = n] \times P^\omega(X(\omega, T_N) \in dy | X(\omega, 0) = x)$$

$$= E^\omega\left[(1 - X(\omega, T_N))^{R_T(\omega,(T^-)-)} | R_T(\omega, 0-) = n, X(\omega, 0) = x \right].$$

Partitioning first on the values of $R_T(\omega, (T - T_N)-)$ and then on the values of $X(\omega, T_N-)$ and using Lemma 6.2 we get that the previous expression equals

$$\sum_{k \in \mathbb{N}_0} E^\omega ((1 - X(\omega, T_N))^k | X(\omega, 0) = x] \times P^\omega(R_T(\omega, (T - T_N)-) = k | R_T(\omega, 0-) = n)$$

$$= \sum_{k \in \mathbb{N}_0} \int_0^1 E^\omega ((1 - X(\omega, T_N))^k | X(\omega, T_N-) = y | X(\omega, 0) = x]$$

$$\times P^\omega(R_T(\omega, (T - T_N)-) = k | R_T(\omega, 0-) = n)$$

$$= \sum_{k \in \mathbb{N}_0} \int_0^1 E^\omega ((1 - y)^{R_T(\omega, T_N-)} | R_T(\omega, (T - T_N)-) = k] \times P^\omega(X(\omega, T_N-) \in dy | X(\omega, 0) = x)$$

$$\times P^\omega(R_T(\omega, (T - T_N)-) = k | R_T(\omega, 0-) = n)$$

$$= E^\omega\left[(1 - X(\omega, T_N-))^{R_T(\omega,(T^-)-)} | R_T(\omega, 0-) = n, X(\omega, 0) = x \right].$$

Iterating this procedure, using successively Lemma 6.1 and Lemma 6.2 (the first one is applied on the intervals $(T_i, T_{i+1})$, while the second one is applied at the times $(T_i)_{i=1}^N$), we finally obtain that

$$E^\omega[(1 - X(\omega, T))^n | X(\omega, 0) = x]$$

equals

$$E^\omega\left[(1 - X(\omega, 0))^{R_T(\omega,(T^-)-)} | R_T(\omega, 0-) = n, X(\omega, 0) = x \right] = E^\omega\left[(1 - x)^{R_T(\omega,(T^-)-)} | R_T(\omega, 0-) = n \right].$$

achieving the proof. \hfill \Box

Proof of Theorem 2.12 (Quenched type frequency from the far past). Applying Theorem 2.11 between $-T$ and 0, we obtain

$$E^\omega [(1 - X(\omega, 0))^n | X(\omega, -T) = x] = E^\omega\left[(1 - x)^{R_0(\omega,T^-)} | R_0(\omega, 0-) = n \right].$$

(6.3)

Since we assume that $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$, the right hand side converges to $W_n(\omega)$, which proves that the moment of order $n$ of $1 - X(\omega, 0)$ conditionally on $(X(\omega, T))^n$ converges to $W_n(\omega)$. Since we are dealing with random variables supported on $[0, 1]$, the convergence of the positive entire moments proves the convergence in distribution and the fact that the limit distribution $\mathcal{L}$ satisfies (2.8).

It remains to prove (2.9). If $\mu$ is a probability measure on $\mathbb{N}_0$ with finite support, let $\tilde{\mu}(\omega)$ denote the distribution of $R_0(\omega, s-) > 0$ given that $R_0(\omega, 0-) \sim \mu$. Note that as far as $R_0(\omega, s)$ is not absorbed, it will either get absorbed at $\uparrow$ at a rate that is greater than or equal to $\eta \nu_0 > 0$, due to the appearance of a mutation of type 0, or go to 0 before such a mutation occurs. As a consequence $T_{0,\uparrow}$, the absorption time of $R_0(\omega, s)$ at $\{0, \uparrow\}$, is stochastically bounded by an exponential random variable with parameter $\eta \nu_0$. Therefore,

$$\tilde{\mu}(T) = P_{\mu}(R_0(\omega, T^-) \in \mathbb{N}) = E^\omega_{\tilde{\mu}}(T_{0,\uparrow} > T) \leq e^{-\eta \nu_0 T}.$$
We thus get that
\[ \tilde{\mu}_T(\omega)(0) \leq \mathbb{E}_\mu^\omega \left[ (1-x)R_0(\omega,T^-) \right] = \tilde{\mu}_T(\omega)(0) + \sum_{k \geq 1} (1-x)k \tilde{\mu}_T(\omega)(k) \leq \tilde{\mu}_T(\omega)(0) + e^{-\theta_0 n_0 T}. \] (6.4)

Similarly, we have
\[ \mathbb{E}_\mu^\omega \left[ (1-x)R_0(\omega,T^-) \right] = \tilde{\mu}_T(\omega)(0) + \sum_{k \geq 1} (1-x)k \tilde{\mu}_T(\omega)(k) \leq \tilde{\mu}_T(\omega)(0) + e^{-\theta_0 n_0 T}, \]
so
\[ \tilde{\mu}_T(\omega)(0) \leq \mathbb{E}_\mu^\omega \left[ (1-x)^{R_0(\omega,T^-)} - 1 \right] \leq \tilde{\mu}_T(\omega)(0) + e^{-\theta_0 n_0 T}. \] (6.5)

Recall from Subsection 2.5.2 that \( W_n(\omega) := \mathbb{P}^\omega(\exists s \geq 0 \text{ s.t. } R_0(\omega,s) = 0 \mid R_0(\omega,0-) = n). \) Choosing \( \mu = \delta_n \) in (6.4) and in (6.5) and subtracting both inequalities we get
\[ \mathbb{E}_\mu^\omega \left[ (1-x)^{R_0(\omega,T^-)} - 1 \right] = W_n(\omega) \geq \mathbb{E}_\mu^\omega \left[ (1-x)^{R_0(\omega,T^-)} \right] - W_n(\omega) \leq e^{-\theta_0 n_0 T}. \]

This inequality together with (6.4) yields the desired result. \( \square \)

**Proof of Lemma 2.5.4** For any \( i \in \mathbb{N}_0^1 \), let \( e_i \) be the vector defined via \( e_i \) for \( j \neq i \). Let us order \( N^0_1 \) as \{1, 0, 1, 2, \ldots\}. The matrix \((Q^0)\) is upper triangular with diagonal elements \(-\lambda_1, -\lambda_0, -\lambda_1, -\lambda_2, \ldots\). For any \( n \in \mathbb{N}_0^1 \), the vector \( \nu_n \) is the eigenvector of \((Q^0)\) associated with the eigenvalue \(-\lambda_n\) and for which the coordinate with respect to \( e_1 \) is 1. It is not difficult to see that these eigenvectors exist and that we have \( v_1 = e_1 \) and \( v_0 = 0 \). For \( n \geq 1 \), writing \( v_n = e_n + e_{n-1}v_{n-1} + \ldots + e_0v_0 + e_1 \) and multiplying by \( e_\lambda(\tilde{Q}^0)\) on both sides, we obtain another expression of \( v_n \) as a linear combination of \( e_n, e_{n-1}, \ldots, e_0, e_1 \). Hence, identifying both expressions, we get an expression for \( e_{n-1} \) and, for each \( k \leq n-2 \), for the coefficient \( c_k \) in term of \( c_{k+1}, \ldots, c_{n-1} \). This allows to see that these coefficients are equal to the coefficients \( v_{n,k} \) defined in the statement of the lemma. More precisely,
\[ v_n = v_{n,n}e_n + v_{n,n-1}e_{n-1} + \ldots + v_{n,0}e_0 + v_{n,n}e_1. \]
Similarly we can show that
\[ e_n = u_{n,n}v_n + u_{n,n-1}v_{n-1} + \ldots + u_{n,0}v_0 + u_{n,n}v_1. \]

We thus get that \((V^\top U^\top = U^\top V^\top = 1d)\) and \((Q^0)\) is diagonal defined in the statement of the lemma and where the matrix products are well-defined since they involve sums of finitely many non-zero terms. The result follows. \( \square \)

**Proof of Theorem 2.7.4** We assume that \( \sigma = 0 \), \( \theta > 0 \) and \( \nu_0, \nu_1 \in (0,1) \). Let us first justify that the matrix products in (2.13), (2.14) and (2.16) are well-defined and that \( C^\top_{i,k}(\omega,T) = 0 \) for all \( k > n2^N \).

Note that if we order \( N^0_1 \) as \{1, 0, 1, 2, \ldots\} then the matrices \( U^\top \) and \( V^\top \) are upper triangular. Moreover, \( B_j(z) = 0 \) for \( i > 2j \). Therefore, for any \( n \in \mathbb{N} \) and any vector \( v = (v_i)_{i \in \mathbb{N}_0^1} \) such that \( v_0 = 0 \) for all \( i > n \), \( v := \sum_{i \geq 0} \beta^i B(z) \sum_{i \geq 0} (V^\top)^i v \) is well-defined and satisfies \( \tilde{v}_i = 0 \) for all \( i > 2n \). In particular \( \beta^i B(z) \sum_{i \geq 0} (V^\top)^i v \) is well-defined. Since \((T_{m-1} - T_m)D_1\) is diagonal, we have that for any \( m \geq 1 \), the product defining the matrix \( A^m(\omega) \) in (2.13) is well-defined. Moreover, for any \( n \in \mathbb{N} \) and any vector \( v = (v_i)_{i \in \mathbb{N}_0^1} \) such that \( v_i = 0 \) for all \( i > n \), the vector \( \tilde{v} := \sum_{i \geq 0} \beta^i e_i(\omega)v \) satisfies \( \tilde{v}_i = 0 \) for all \( i > 2n \). In particular, for any \( m \geq 1 \), the product \( \exp(-(T_N + T)D_1)A^m(\omega)A^{m-1}(\omega) \ldots A^1(\omega)U^\top \) is well-defined. Additionally, for \( n \geq 1 \) and a vector \( v = (v_i)_{i \in \mathbb{N}_0^1} \) such that \( v_i = 0 \) for all \( i > n \), the vector \( \tilde{v} := \exp(-(T_N + T)D_1)A^m(\omega)A^{m-1}(\omega) \ldots A^1(\omega)U^\top \) satisfies \( \tilde{v}_i = 0 \) for all \( i > 2m \). Transposing, we see that the matrix \( C^\top(\omega,T) \) in (2.16) is well-defined and satisfies \( C^\top_{i,k}(\omega,T) = 0 \) for all \( k > n2^N \).

Now, for \( s > 0 \), we define the stochastic matrix \( P^s_\mu(\omega) := (p_{i,j}^s(\omega,s) = j \mid R_0(\omega,0-) = i) \) via
\[ p_{i,j}^s(\omega,s) := \mathbb{P}^\omega(R_0(\omega,s) = j \mid R_0(\omega,0-) = i). \]
Hence, defining $\rho(y) := (y^j)_{j \in \mathbb{N}}^+$, $y \in [0, 1]$ (with the convention $y^+ := 0$), we obtain
\[
\mathbb{E}^\omega[y^{R_0(\omega, T^-)}] \mid R_0(\omega, T^-) = n = (\mathbb{P}_T^1(\omega)\rho(y))_n = (\mathbb{P}_T(\omega)U_1 P_1(y))_n, \tag{6.6}
\]
where we have used that $\rho(y) = U_1 P_1(y)$ with $P_1(y) = (P^1_k(y))_{k \in \mathbb{N}}$, Thus, Theorem\textsuperscript{[2.13]} and Eq. (6.6) yield
\[
\mathbb{E}^\omega[(1 - X(\omega, 0))^n] \mid X(\omega, -T^-) = x = \sum_{k=0}^{\infty} (\mathbb{P}_T^1(\omega)U_1)_{n,k} P^1_k(1 - x). \tag{6.7}
\]
Now, consider the semi-group $M_t := (M_t(s))_{s \geq 0}$ of the killed ASG in the null environment, which is defined via $M_t(s) := \exp(sQ^1_t)$. Thanks to Lemma\textsuperscript{[2.13]} $M_t(\beta) = U_1 E^1_t(\beta)V_t$, where $E^1_t(\beta)$ is the diagonal matrix with diagonal entries $(e^{-\lambda^1_j})_{j \in \mathbb{N}}$. We split the proof the first statement in two cases:

**Case 1, when $\omega$ has no jumps in $[-T, 0]$:** In this case, we have
\[
\mathbb{P}_T^1(\omega)U_1 = M_t(T)U_1 = U_1 E^1_t(\beta)V_tU_1 = U_1 E^1_t(T),
\]
where in the last identity we used that $U_1 V_1 = Id$. Since the right hand side in the previous identity coincides with $C^1(\omega, T)$, the proof of the first part of the statement follows from (6.7).

**Case 2, when $\omega$ has at least one jump in $[-T, 0]$:** In this case, let $T_N < T_{N-1} < \cdots < T_i$ denote the sequence of jumping times of $\omega$ in $[-T, 0]$. Disintegrating on the values of $R_0(\omega, -(T_i-))$ and $R_0(\omega, -T_i)$, $i \in [N]$, we see that
\[
\mathbb{P}_T^1(\omega) = M_t(-T_i)B(\Delta(\omega(T_i)))M_t(T_i - T_{i-2})B(\Delta(\omega(T_{i-2}))) \cdots B(\Delta(\omega(T_N)))M_t(T_N + T). \tag{6.8}
\]
Using this, the relation $M_t(\beta) = U_1 E^1_t(\beta)V_t$, the definition of the matrices $\beta^\dagger$ and $A^\dagger_t$ (see (2.13) and (2.14) resp.), and the fact that $U_1 V_1 = Id$, we obtain
\[
\mathbb{P}_T^1(\omega)U_1 = U_1 E^1_t(-T_i)\beta^\dagger(\Delta(\omega(T_i))\dagger E^1_t(T_i - T_{i-2})\beta^\dagger(\Delta(\omega(T_{i-2})) \cdots \beta^\dagger(\Delta(\omega(T_N))\dagger E^1_t(T_N + T)
\]
\[
= U_1 A^\dagger_t(\omega)\dagger A^\dagger_t(\omega)\dagger \cdots A^\dagger_t(\beta^\dagger(\Delta(\omega(T_N))\dagger E^1_t(T_N + T)
\]
\[
= C(\omega, T),
\]
proving the first part of the statement in this case.

It remains to prove that $C^\dagger_{n,0}(\omega, T)$ converges to $W_n(\omega)$ as $T$ tends to infinity. In the case of the null environment, i.e. $\omega = 0$, the first part of the statement together with (6.6) yield
\[
\mathbb{E}^\omega[y^{R_0(0, T^-)}] \mid R_0(0, 0^-) = n = \sum_{k=0}^{n} e^{-\lambda^1_k T} w^\dagger_{n,k} P^1_k(y).
\]
Since $\lambda^1_k > 0$ for $k \in \mathbb{N}$ and $\lambda^1_0 = 0$, the proof of the second part of the statement follows by letting $T \to \infty$ in the previous identity.

The general case is a direct consequence of the following proposition.

**Proposition 6.3.** Assume that $\sigma > 0$, $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$, then we have
\[
\left| C^\dagger_{n,0}(\omega, T) - W_n(\omega) \right| \leq e^{-\theta \nu_0 T}.
\]

**Proof.** Let $\omega_T$ be the environment that coincides with $\omega$ in $(-T, 0]$ and that is constant and equal to $\omega$ in $(-\infty, -T]$. Since $\mathbb{P}_T^1(\omega_T) = \mathbb{P}_T^1(\omega)$ and $\omega_T$ has no jumps in $(-\infty, T]$, we obtain
\[
W_n(\omega_T) = \sum_{k \geq 0} \mathbb{P}_T^1(\omega_T, T) p^\dagger_{n,k}(\omega_T, T) \mathbb{P}^\dagger_{n,k}(\omega_T, T) = \mathbb{P}_T^1(\omega_T, T) W_n(0) = C^\dagger_{n,0}(\omega, T),
\]
where in the last line we used Theorem\textsuperscript{[2.13]} for the null environment (which was already proved) and the definition of $C(\omega, T)$. Now combining (6.9) with (6.4) applied to $\omega_T$ with $\mu = \delta_n$ yields
\[
\mathbb{P}_T^1(\omega_T, T) \leq C^\dagger_{n,0}(\omega, T) \leq \mathbb{P}_T^1(\omega_T, T) e^{-\theta \nu_0 T} = p^\dagger_{n,0}(\omega, T) + e^{-\theta \nu_0 T}.
\]
\[
(6.10)
\]
Then, combining (6.4) applied to \( \omega \) with \( \mu = \delta_n \) and (6.10) we get
\[
C_{n,0}(\omega, T) - e^{-\theta \nu_0 T} \leq W_n(0) \leq C_{n,0}(\omega, T) + e^{-\theta \nu_0 T},
\]
and the result follows. \( \square \)

6.2. Ancestral type distribution: The pruned look-down ASG. In this section, we show the results presented in Sections 6.2.6.3

Proof of Proposition 6.22. The proof is completely analogous to the proof of Proposition 6.2.7 \( \square \)

If \( \mu \) is a probability measure on \( \mathbb{N} \), let \( \mu^{T}_{\beta}(\omega) \) denote the distribution of \( L_T(\omega, \beta-) \) when the initial value \( L_T(\omega, 0-) \) follows distribution \( \mu \).

Proof of Theorem 6.21. Let us first assume that \( \theta > 0 \) and \( \nu_0 > 0 \). We will first show the convergence of \( \mu^{T}_{\beta}(\omega) \) to a limit \( \mu^{\infty}_{\omega}(\omega) \) when \( T \to \infty \), which does not depend on the choice of \( \mu \).

We first let \( t_2 > t_1 > 0 \) and study \( d_{TV}(\mu^{T}_{t_2}(\omega), \mu^{T}_{t_1}(\omega)) \). Note that we have
\[
d_{TV}(\mu^{T}_{t_2}(\omega), \mu^{T}_{t_1}(\omega)) = d_{TV}((\mu^{T}_{i_2-t_1}(\omega))^{T}_{i_1}(\omega), \mu^{T}_{i_1}(\omega)).
\]

Thus, we are led to study \( d_{TV}(\mu^{T}_{i_2}(\omega), \mu^{T}_{i_1}(\omega)) \), i.e. the total variation distance between the distributions of \( L_t(\omega, t-) \) with two different starting laws \( \mu \) and \( \tilde{\mu} \).

From the dynamic of \( L_T(\omega, \cdot) \), we know that the transition from any from any state \( i \) to the state 1 is \( q^0(i, 1) \), which is at least \( \theta \nu_0 > 0 \). Let \( \tilde{L}^{T}_{\beta}(\omega, \cdot) \) be a process with initial distribution \( \mu \) and with the same dynamic as \( L_T(\omega, \cdot) \), excepting by the transition rate to the state 1, which is \( q^0(i, 1) - \nu_0 \theta > 0 \). We decompose the dynamic of the \( L_t(\omega, \cdot) \) with starting law \( \mu \) as follows: (1) \( L_t(\omega, \cdot) \) has the same dynamic as \( \tilde{L}^{T}_{\beta}(\omega, \cdot) \) on \([0, \xi] \), where \( \xi \) is an independent exponential random variable with parameter \( \theta \nu_0 \), (2) at time \( \xi \) the process jumps to the state 1 regardless to its current position, and (3) conditionally on \( \xi \), \( L(\omega, \cdot) \) has the same law on \([ \xi, \infty) \) as an independent copy of \( L_{T-\xi}(\omega, \cdot) \) started with one line.

Using this idea, we couple, on the basis of the same random variable \( \xi \), two copies \( L_T(\omega, \cdot) \) and \( \tilde{L}^{T}_{\beta}(\omega, \cdot) \) of \( L_T(\omega, \cdot) \) with starting laws \( \mu \) and \( \tilde{\mu} \), respectively, so that the two processes are equal on \([ \xi, \infty) \). Since \( \tilde{L}(\omega, T-\xi) \sim \mu^{T}_{\beta}(\omega) \) and \( \tilde{\xi}(\omega, T-\xi) \sim \mu^{T}_{\beta}(\omega) \), we have
\[
d_{TV}(\mu^{T}_{\beta}(\omega), \mu^{T}_{\beta}(\omega)) \leq P \left( \tilde{L}(\omega, T-\xi) \neq \tilde{\xi}(\omega, T-\xi) \right) \leq P(\xi > T) = e^{-\theta \nu_0 T}.
\]

Putting into (6.11) we obtain, for any starting law \( \mu \) and any \( t_2 > t_1 > 0 \),
\[
d_{TV}(\mu^{T}_{i_2}(\omega), \mu^{T}_{i_1}(\omega)) \leq e^{-\theta \nu_0 t_1}.
\]

In particular \( (\mu^{T}_{\beta}(\omega))_{t>0} \) is Cauchy as \( t \to \infty \) for the total-variation distance and therefore convergent. Hence, \( \mu^{\infty}_{\omega}(\omega) \) is well-defined for any starting law \( \mu \). Moreover, (6.12) implies that \( \mu^{\infty}_{\omega}(\omega) \) does not depend on \( \mu \). Therefore, the first claim in Theorem 6.21 is proved.

Setting \( t_1 = T \) and letting \( t_2 \to \infty \) in (6.13) yields
\[
d_{TV}(\mu^{\infty}_{\omega}(\omega), \mu^{T}_{\beta}(\omega)) \leq e^{-\theta \nu_0 T}.
\]

Since \( h^{\omega}(x) = 1 - E\left[(1-x)^{Z^{\omega}(\omega)}\right] \) and \( h^{\omega}_T(x) = 1 - E\left[(1-x)^{Z^{\omega}(\omega)}\right] \), where \( Z^{\omega}(\omega) \sim (\delta_1)^{\infty}_{\omega}(\omega) \) and \( Z^{\omega}_T(\omega) \sim (\delta_1)^{\infty}_{T_0}(\omega) \), we get
\[
|h^{\omega}(x) - h^{\omega}_T(x)| \leq d_{TV}((\delta_1)^{\infty}_{\omega}(\omega), (\delta_1)^{\infty}_T(\omega)) \leq e^{-\theta \nu_0 T},
\]
achieving the proof in the case \( \theta > 0 \) and \( \nu_0 > 0 \).

Let us now assume that \( \theta \nu_0 = 0 \). As before, the proof will be based on an appropriate upper bound for \( d_{TV}(\mu^{T}_{\beta}(\omega), \mu^{T}_{\beta}(\omega)) \).

Let \( 0 < T_1 < T_2 < \cdots \) denote the sequence of the jumping times of \( \omega \) and set \( T_0 := 0 \) for convenience. On each time interval \((T_i, T_{i+1})\), \( L_T(\omega, \cdot) \) has transition rates given by \( (q^0(i, j))_{i,j \in \mathbb{N}} \). For any \( k > l \), let \( H(k, l) \) denote the hitting time of \([l] \) by a Markov chain starting at \( k \) and with transition rates given by \( (q^0(i, j))_{i,j \in \mathbb{N}} \). Let \( (S_i)_{i \geq 2} \) be a sequence of independent random variables such that for each \( l \geq 2 \),

\[
C_{n,0}(\omega, T) - e^{-\theta \nu_0 T} \leq W_n(0) \leq C_{n,0}(\omega, T) + e^{-\theta \nu_0 T},
\]
$S_i \sim H(l, l - 1)$. By the Markov property we have that for any $k \geq 1$, $H_k(1, 1)$ is stochastically upper bounded by $\sum_{i=2}^{k} S_i$. Therefore, for any $i$ such that $T_{i+1} < T$ and any $k \geq 1$ we have

$$\mathbb{P}(L_T(\omega, (T - T_i)^-) = 1 | L_T(\omega, (T - T_{i+1}) = k) \geq \mathbb{P}\left( \sum_{i=2}^{k} S_i \leq T_{i+1} - T_i \right) \geq \mathbb{P}\left( \sum_{i=2}^{\infty} S_i \leq T_{i+1} - T_i \right).$$

Let $l_0$ be such that $\sigma l \leq (l - 1)/4$ for all $l \geq l_0$. For $l \geq l_0$, we define a random walk $Z^l := (Z^l(t))_{t \geq 0}$ on $\{1, 2, 3, 4\}$ starting at $l$ and such that:

1. It jumps from $n \geq l$ at rate $(n - l)n$ to either $n - 1$ or $n + 1$, with probability $3/4$ and $1/4$ respectively,

2. It is absorbed at $l - 1$. Let $W_l$ denote the hitting time of $l - 1$ by $Z^l$. We can see that $\mathbb{E}[W_l] \leq 2/(l - 1)$.

Note that $Y^l$ always jumps to the right with a smaller rate than $Z^l$ and jumps to the left at a higher rate. Moreover, the jumps of both are always of size 1. Therefore, it is possible to build a coupling of $Y^l$ and $Z^l$ such that for every $t \geq 0, Y^l(t) \leq Z^l(t)$. In this coupling the hitting time of $l - 1$ by $Y^l$ (which is almost equal in law to $S_l$) is almost surely smaller or equal to $W_l$. We thus get, for all $l \geq l_0$

$$\mathbb{E}[S_l] \leq \mathbb{E}[W_l] \leq \frac{2}{(l - 1)}.$$

Therefore, $\sum_{i=0}^{\infty} \mathbb{E}[S_l] < +\infty$. Thus, $S^\infty := \sum_{i=2}^{\infty} S_l$ is almost surely finite. Moreover, since for each $l \geq 2$, the support of $S_l$ contains 0, the support of $S^\infty$ contains 0. In particular $q(s) := \mathbb{P}(\sum_{i=0}^{\infty} S_l \leq s)$ is positive for all $s > 0$. From above we get

$$\forall k \geq 1, \quad \mathbb{P}(L_T(\omega, (T - T_i)^-) = 1 | L_T(\omega, (T - T_{i+1}) = k) \geq q(T_{i+1} - T_i).$$

Let us consider $(V_T(\omega, \beta))_{\beta \geq 0}$ and $(\tilde{V}_T(\omega, \beta))_{\beta \geq 0}$, two independent versions of $(L_T(\omega, \beta))_{\beta \geq 0}$ with starting laws respectively $\mu$ and $\tilde{\mu}$ at instant $\beta = 0$. For any $i \geq 1$, let $(V_T(\omega, \beta))_{\beta \geq 0}$ be a version of $(L_T(\omega, \beta))_{\beta \geq 0}$ starting at $1$ at instant $\beta = (T - T_i)^-$, independent from $(V_T(\omega, \beta))_{\beta \geq 0}$ and $(\tilde{V}_T(\omega, \beta))_{\beta \geq 0}$.

Let $I_0(T)$ be the largest index $i$ such that 1) $T_i < T$ and 2) $V_T(\omega, (T - T_i)^-) = \tilde{V}_T(\omega, (T - T_i)^-) = 1$, if such an index exists, and let $I_0(T) := +1$ otherwise. We define $(U_T(\omega, \beta))_{\beta \geq 0}$ and $(\tilde{U}_T(\omega, \beta))_{\beta \geq 0}$ in the following way:

$$U_T(\omega, \beta) := \begin{cases} V_T(\omega, \beta) & \text{if } \beta < T - T_{I_0(T)}, \\
\tilde{V}_T(\omega, \beta) & \text{if } \beta \geq T - T_{I_0(T)}, \\
\tilde{U}_T(\omega, \beta) & \text{if } \beta > T - T_{I_0(T)}.
\end{cases}$$

Note that $(U_T(\omega, \beta))_{\beta \geq 0}$ and $(\tilde{U}_T(\omega, \beta))_{\beta \geq 0}$ are realizations of $(L_T(\omega, \beta))_{\beta \geq 0}$ with starting laws respectively $\mu$ and $\tilde{\mu}$ at instant $\beta = 0$ (in particular we have $U_T(\omega, T^-) \sim \mu_T^1(\omega, \tilde{\mu}_T^1(\omega)$).

Moreover we have $U_T(\omega, \beta) = \tilde{U}_T(\omega, \beta)$ for all $\beta \geq T - T_{I_0(T)}$. Therefore,

$$d_{TV}(\mu_T^1(\omega), \tilde{\mu}_T^1(\omega)) \leq \mathbb{P}\left( V_T(\omega, T^-) \neq \tilde{U}_T(\omega, T^-) \right) \leq \mathbb{P}(I_0(T) = 1).$$

Let $N(T)$ be the index of the last jumping time before $T$ (so that $T_1 < T_2 < \cdots < T_{N(T)}$ are the jumping times of $\omega$ on $[0, T]$). According to (6.13), we have that for any $k_1, k_2 \geq 1$ with $k_1 \neq k_2$,

$$\mathbb{P}\left( V_T(\omega, (T - T_i)^-) \neq 1 \text{ or } \tilde{V}_T(\omega, (T - T_i)^-) \neq 1 \mid V_T(\omega, (T - T_{i+1}) = k_1, \tilde{V}_T(\omega, (T - T_{i+1}) = k_2 \right) \leq 1 - q(T_{i+1} - T_i)^2.$$

Therefore,

$$\mathbb{P}(I_0(T) = 1) \leq \prod_{i=1}^{N(T)} (1 - q(T_i - T_{i-1})^2) =: \varphi_\omega(T).$$

Putting in (6.13) we deduce that $d_{TV}(\mu_T^1(\omega), \tilde{\mu}_T^1(\omega)) \leq \varphi_\omega(T)$. Note that $\varphi_\omega$ does not depend on the particular choice of $\mu$ and $\tilde{\mu}$. Recall that by assumption the sequence of jumping times $T_1, T_2, \ldots$ is infinite and the distance between the successive jumps does not converge to 0. Therefore there is $\epsilon > 0$ such that for infinitely many indices $i$ we have $T_{i+1} - T_i > \epsilon$. The number of factors smaller than $1 - q(\epsilon)^2 < 1$ in the product defining $\varphi_\omega(T)$ thus converges to infinity when $T$ goes to infinity. We deduce that $\varphi_\omega(T) \to 0$ as $T \to \infty$, which concludes the proof. 

□
Proof of Theorem 2.26. We are interested in the generating function of $L_T(\omega, T^-)$. For $s > 0$, we define the stochastic matrix $\mathcal{P}_s^T(\omega) := (p_{i,j}^T(\omega, s))_{i,j \in \mathbb{N}}$ via

$$p_{i,j}^T(\omega, s) := \mathbb{P}(L_T(\omega, s^-) = j \mid L_T(\omega, 0^-) = i).$$

We also define $\{M(s)\}_{s \geq 0}$ via $M(s) := \exp(sQ^T)$, i.e., $M$ is the semi-group of the pruned LD-ASG in the null environment. Let $T_1 < T_2 < \cdots < T_N$ denote the sequence of jumping times of $\omega$ in $[0, T]$. Then, disintegrating on the values of $L_T(\omega, (T-T_1))$ and $L_T(\omega, T-T_i)$, $i \in [N]$, we see that

$$\mathcal{P}_0^T(\omega) = M(T-T_N)B(\Delta \omega(T_N))M(T_N-T_{N-1})B(\Delta \omega(T_{N-1})) \cdots B(\Delta \omega(T_1))M(T_1).$$

(6.16)

In addition,

$$\mathbb{E}^\omega[y^{L_T(\omega, T^-)} \mid L_T(\omega, 0^-) = n] = (\mathcal{P}_0^T(\omega)\rho(y))_n, \quad \text{where} \quad \rho(y) := (y^i)_{i \in \mathbb{N}}.$$

(6.17)

Thanks to Lemma 2.22 we have that $M(\beta) = UE(\beta)V$, where $E(\beta)$ is the diagonal matrix with diagonal entries $(e^{-\lambda_i\beta})_{i \in \mathbb{N}}$. Moreover, using this, Eq. (6.16), the relation $M(\beta) = UE(\beta)V$, the definition of the matrices $\beta()$ in 2.29, the fact that $UV = Id$, and the definition of the matrices $A$ in 2.29, we obtain

$$\mathcal{P}_0^T(\omega)\rho(y) = U E(T-T_N)A_N(\omega)^\top A_N-1(\omega)^\top \cdots A_1(\omega)^\top P(y)$$

$$= U E(T-T_N)A_N(\omega)^\top A_N-1(\omega)^\top \cdots A_1(\omega)^\top P(y)$$

$$= U E(T-T_N)[A_1(\omega)A_2(\omega) \cdots A_N(\omega)]^\top P(y).$$

(6.18)

Now using the previous identity, Proposition 2.20 and Eq. (6.17), we obtain

$$h_{T}^\omega(x) = 1 - \mathbb{E}^\omega[(1-x)L_T(\omega, T^-) \mid L_T(\omega, 0^-) = 1] = 1 - \sum_{k=0}^{\infty} C_{1,k}(\omega, T) P_k(1-x).$$

Similarly as in the proof of Theorem 2.14, we can show that $C_{1,k}(\omega, T) = 0$ for all $k > 2^N$. (2.30) follows. Let us now show that for each $k \geq 1$, the coefficient $C_{1,k}(\omega, T)$ converges to a real limit when $t$ goes to infinity. We see from (2.30) that the decomposition of the generating function of $L_T(\omega, T^-)$ (conditionally on $L_T(\omega, 0^-) = 1$) in the basis of polynomials $(P_k(y))_{k \in \mathbb{N}}$ is given by $\sum_{k=1}^{\infty} C_{1,k}(\omega, T) P_k(y)$. In the basis $(y^k)_{k \in \mathbb{N}}$, this decomposition is clearly given by $\sum_{k=1}^{\infty} \mathbb{P}(L_T(\omega, T^-) = k \mid L_T(\omega, 0^-) = 1) y^k$. Since $U^\top$ is the transition matrix from the basis $(y^k)_{k \in \mathbb{N}}$ to the basis $(P_k(y))_{k \in \mathbb{N}}$ we deduce that

$$C_{1,k}(\omega, T) = \sum_{i \in \mathbb{N}} u_{i,k} \mathbb{P}(L_T(\omega, T^-) = i \mid L_T(\omega, 0^-) = 1),$$

or equivalently,

$$C_{1,k}(\omega, T) = \mathbb{E}[u_{L_T(\omega, T^-),k} L_T(\omega, 0^-) = 1].$$

From Theorem 2.23 we know that the distribution of $L_T(\omega, T^-)$ converges when $T \to \infty$. In addition, Lemma 6.3 tells us that the function $i \mapsto u_{i,k}$ is bounded, and hence $C_{1,k}(\omega, T)$ converges to a real limit. Recall that $T_1 < T_2 < \cdots$ is the increasing sequence of the jumping times of $\omega$ and that this sequence converges to infinity. Therefore

$$\lim_{T \to \infty} C_{1,k}(\omega, T) = \lim_{m \to \infty} C_{1,k}(\omega, T_m) = \lim_{m \to \infty} U [A_1(\omega)A_2(\omega) \cdots A_N(\omega)]^\top,$$

where we have used (2.31) for the last equality. This shows in particular that the limit in the right hand side of (2.33) exists and equals $\lim_{T \to \infty} C_{1,k}(\omega, T)$.

We now prove (2.32) together with the convergence of the series in this expression. We already know from Theorem 2.23 that $h_{T}^\omega(x)$ converges to $h^\omega(x)$ when $T \to \infty$ and we have just proved that for any $k \geq 1$, $C_{1,k}(\omega, T)$ converges to $C_{1,k}(\omega, \infty)$, defined in (2.33), when $T \to \infty$. In addition, we claim that for all $y \in [0,1]$ and $T > T_1$, we have

$$|C_{1,k}(\omega, T)P_k(y)| \leq 4^k \times (2ek)^{(k+\theta)/2} e^{-\lambda_k T_1}.$$  

(6.19)

Using this, the expression (2.32) for $h^\omega(x)$ and the convergence of the series would follow using the dominated convergence theorem. It only remains to prove (6.19). We saw above that the decomposition of the generating function of $L_T(\omega, T^-)$ (conditionally on $L_T(\omega, 0^-) = 1$) in the basis of polynomials $(P_k(y))_{k \in \mathbb{N}}$
Lemma 6.4. For all $k \geq 1$, the decomposition of the generating function of $L^T(\omega, T - T_1)$ (conditionally on $L^T(\omega, 0^+) = 1$) in the basis of polynomials $(P_k(y))_{k \in \mathbb{N}}$ can be expressed as $\sum_{k=1}^{\infty} \hat{C}_{1,k}(\omega, T) P_k(y)$, where $\hat{C}_{1,k}(\omega, T) = \mathbb{E}[u_{L^T(\omega, T - T_1), k} | L^T(\omega, 0^+) = 1]$. Similarly we can see that for $T > T_1$, this together with Lemma 6.4 imply that, for all $\lambda$ of $\mathbb{R}$, we have used Lemma 6.5, stated just after the proof, for the last inequality. Since, by the definition of $\sum_{k=1}^{\infty} \hat{C}_{1,k}(\omega, T) P_k(y)$, where $\hat{C}_{1,k}(\omega, T) = \mathbb{E}[u_{L^T(\omega, T - T_1), k} | L^T(\omega, 0^+) = 1]$. Similarly as we proved (6.18), we can prove that

$$P_{T_1+k}(\omega) = U E(T - T_N) \beta(\Delta \omega(T_N))^T E(T_N - T_{N-1}) \beta(\Delta \omega(T_{N-1})) \cdots \beta(\Delta \omega(T_1))^T P(y)$$

(6.20)

Since $E(T_1)$ is the diagonal matrix with diagonal entries $(e^{-\lambda T_1})_{\omega \in \mathbb{N}}$ we conclude from (6.19) and (6.20) that $C_{1,k}(\omega, T) = e^{-\lambda_k T_1} \hat{C}_{1,k}(\omega, T)$. Therefore

$$C_{1,k}(\omega, T) = e^{-\lambda_k T_1} \mathbb{E}[u_{L^T(\omega, T - T_1), k} | L^T(\omega, 0^+) = 1].$$

This together with Lemma 6.4 imply that, for all $k \geq 1$ and $t \geq 0$,

$$|C_{1,k}(\omega, T)| \leq (2\epsilon k)^{(k+\theta)/2} e^{-\lambda_k T_1}.$$

Combing this with Lemma 6.6 we obtain (6.10), which concludes the proof. \hfill \Box

Lemma 6.4. For all $k \geq 1$

$$\sup_{j \geq 1} u_{j,k} \leq (2k)^{(k+\theta)/2}.$$ 

Proof. Let $k \geq 1$. By the definition of the matrix $U$ in Lemma 2.22 the sequence $(u_{j,k})_{j \geq 1}$ satisfies

$$u_{j,k} = 0 \text{ for } j < k, \quad u_{k,k} = 1, \quad u_{k+1,k} = \frac{\gamma_{k+1}}{\lambda_{k+1} - \lambda_k},$$

$$u_{k+l,k} = \frac{1}{\lambda_{k+l} - \lambda_k} \left( \gamma_{k+l} u_{k+l-1,k} + \nu_0 \theta \left( \sum_{j=0}^{l-2} u_{k+j,k} \right) \right) \quad \text{for } l \geq 2.$$ 

Let $M_k^l := \sup_{1 \leq j} u_{i,k}$. By the definitions of $\gamma_{j+1}$, $\lambda_{k+1}$, $\lambda_k$ in Lemma 2.22 we see that

$$\gamma_{k+1} = \lambda_{k+1} - (k-1)\nu_0 \theta > \lambda_{k+1} - \lambda_k.$$ 

This together with the above expressions yield that

$$M_k^1 = 1, \quad M_k^{k+1} = \frac{\lambda_{k+1} - (k-1)\nu_0 \theta}{\lambda_{k+1} - \lambda_k} \leq \frac{\lambda_{k+1} - \lambda_k}{\lambda_{k+1} - \lambda_k} = 1 + \frac{\lambda_k}{\lambda_{k+1} - \lambda_k},$$

and for $l \geq 2$,

$$M_k^{k+l} \leq M_k^{k+l-1} \vee \frac{M_k^{k+l-1}}{\lambda_{k+l} - \lambda_k} \left( \gamma_{k+l} + (l-1)\nu_0 \theta \right)$$

$$= M_k^{k+l-1} \vee M_k^{k+l-1} \times \frac{\lambda_{k+l} - (k-1)\nu_0 \theta}{\lambda_{k+l} - \lambda_k}$$

$$= M_k^{k+l-1} \times \frac{\lambda_{k+l} - (k-1)\nu_0 \theta}{\lambda_{k+l} - \lambda_k}$$

$$\leq M_k^{k+l-1} \times \frac{\lambda_{k+l}}{\lambda_{k+l} - \lambda_k}$$

$$\leq M_k^{k+l-1} \times \left( \frac{1}{1 + \frac{\lambda_k}{\lambda_{k+l} - \lambda_k}} \right).$$

As a consequence, we have

$$\sup_{j \geq 1} u_{j,k} \leq \prod_{l=1}^{+\infty} \left( 1 + \frac{\lambda_k}{\lambda_{k+l} - \lambda_k} \right) =: M_k^\infty$$

(6.21)

Since $\lambda_{k+l} \sim l^2$ as $l \to \infty$, it is easy to see that the infinite product $M_k^\infty$ is finite. Then,

$$M_k^\infty = \exp \left[ \sum_{l=1}^{+\infty} \log \left( 1 + \frac{\lambda_k}{\lambda_{k+l} - \lambda_k} \right) \right] \leq \exp \left[ \sum_{l=1}^{+\infty} \frac{\lambda_k}{\lambda_{k+l} - \lambda_k} \right] \leq \exp \left[ \frac{\lambda_k \log(2\epsilon k)}{2(k - 1)} \right],$$

where we have used Lemma 6.5, stated just after the proof, for the last inequality. Since, by the definition of $\lambda_k$ in Lemma 2.22 we have $\lambda_k = (k-1)(k+\theta)$, we obtain the asserted result. \hfill \Box
Lemma 6.5. For all $k \in \mathbb{N}$
\[
\sum_{i=1}^{+\infty} \frac{1}{\lambda_{k+i} - \lambda_k} \leq \log(2ek) \frac{1}{2(k-1)}.
\]

Proof. Using the definition of $\lambda_k$ in Lemma 2.22 we have
\[
\sum_{i=1}^{+\infty} \frac{1}{\lambda_{k+i} - \lambda_k} \leq \sum_{i=1}^{+\infty} \frac{1}{(k+i)(k+i-1) - k(k-1)}
\]
\[
\leq \frac{1}{k+1} + \int_{k}^{+\infty} \frac{1}{x(x+1) - k(k-1)} dx
\]
\[
= \frac{1}{2k} + \int_{1}^{+\infty} \frac{1}{u^2 + (2k-1)u} du = \frac{1}{2k} + \lim_{a \to +\infty} \int_{1}^{a} \frac{1}{u(u+2k-1)} du
\]
\[
= \frac{1}{2k} + \lim_{a \to +\infty} \frac{1}{2k-1} \left( \int_{1}^{a} \frac{1}{u} du - \int_{1}^{a} \frac{1}{u+2k-1} du \right)
\]
\[
= \frac{1}{2k} + \frac{\log(2k)}{2k-1} \leq \frac{1 + \log(2k)}{2(k-1)} = \log(2ek) \frac{1}{2(k-1)}.
\]
\[
\]

Lemma 6.6. For all $k \in \mathbb{N}$, we have
\[
\sup_{y \in [0,1]} |P_k(y)| \leq 4^k.
\]

Proof. By definition of the polynomials $P_k$ in (2.27), we have for $k \geq 1$
\[
\sup_{y \in [0,1]} |P_k(y)| \leq \sum_{i=1}^{k} |v_{k,i}|. \tag{6.22}
\]

Let us fix $k \geq 1$ and define $S^k_j := \sum_{i \geq j} |v_{k,i}|$. Note that $S^k_j = 0$ for $j > k$ and that $S^k_k = 1$ by the definition of the matrix $(v_{k,j})_{i,j \in \mathbb{N}}$ in Lemma 2.22. In particular, the asserted result is true for $k = 1$, we thus assume $k > 1$ from now on. Using (2.26), we see that for any $j \in [k-1]$, $S^k_j = S^k_{j+1} + |v_{k,j}| = S^k_{j+1} + \left( \frac{1}{\lambda_k - \lambda_j} \right) \left( \sum_{i=j+1}^{k} v_{k,i} \right) \left( v_{0,\theta} + v_{k,j+1, \gamma_j+1} \right)$
\[
\leq S^k_{j+1} + \frac{1}{\lambda_k - \lambda_j} \left[ S^k_{j+2} v_{0,\theta} + (S^k_{j+1} - S^k_{j+2}) \gamma_j+1 \right]
\]
\[
\leq \left( 1 + \frac{\gamma_j+1}{\lambda_k - \lambda_j} \right) S^k_{j+1} + \frac{v_{0,\theta} - \gamma_j+1}{\lambda_k - \lambda_j} S^k_{j+2}.
\]

Note that $\frac{v_{0,\theta} - \gamma_j+1}{\lambda_k - \lambda_j} \leq 0$, because of the definition of the coefficients $\gamma_i$ in Lemma 2.22. Thus, for $j \in [k-1]$ $S^k_j \leq \left( 1 + \frac{\gamma_j+1}{\lambda_k - \lambda_j} \right) S^k_{j+1}. \tag{6.23}$

By the definitions of $\gamma_{j+1}$, $\lambda_k$, $\lambda_j$ in Lemma 2.22 we have
\[
\frac{\gamma_{j+1}}{\lambda_k - \lambda_j} = \frac{(j+1)j + j\nu_1 \theta + v_0 \theta}{k(k-1) - j(j-1) + (k-j)\nu_1 \theta + (k-j)v_0 \theta}
\]
\[
\leq \frac{(j+1)j}{k(k-1) - j(j-1)} \left( \frac{j\nu_1 \theta}{(k-j)\nu_1 \theta} + \frac{v_0 \theta}{(k-j)v_0 \theta} \right)
\]
\[
\leq \frac{(j+1)j}{j(k-1) - j(j-1)} \left( \frac{j\nu_1 \theta}{(k-j)\nu_1 \theta} + \frac{v_0 \theta}{(k-j)v_0 \theta} \right) \leq \frac{j+1}{k-j}.
\]

In particular,
\[
1 + \frac{\gamma_{j+1}}{\lambda_k - \lambda_j} \leq \frac{k + j + 2}{k - j}.
\]
Plugging this into (6.29) yields, for all $j \in [k - 1]$

$$S_j^k \leq \frac{k + j + 2}{k - j} S_{j+1}^k.$$ 

Then, applying the above inequality recursively and combining with $S_1^k = 1$ we get

$$\sum_{i=1}^{k} |\eta_{k,i}| = S_k^k = \prod_{j=1}^{k-1} \frac{k + j + 2}{k - j} = \frac{(2k+1)!}{(k-1)!} \leq 2^{2k+1}/2 = 4^k.$$ 

Combining with (6.29) we obtain the asserted result.

\[\square\]

7. An application: exceptional environment in the recent past

Let us now present an application of the methods developed in the present paper that illustrates the interest in studying both quenched and annealed results. More precisely, consider a population living in a stationary random environment, which has been recently subject to a big perturbation. We are interested in the influence of this recent exceptional behavior of the environment, in the absence of knowledge of the environment in the far past.

The process starts at $-\infty$ and we are currently at instant $0$. Recall the the environment is a realization of a Poisson point process $(t_i, p_i)_{i \in I}$ on $(-\infty, 0] \times (0, 1)$ with intensity measure $dt \times \mu$, where $dt$ stands for the Lebesgue measure and $\mu$ is a measure on $(0, 1)$ satisfying $\int x \mu(dx) < \infty$. In this section, we assume that the intensity measure $\mu$ is finite, and hence, the environment is almost surely simple. We moreover assume that the parameter of selection $\sigma$ is zero. We assume that we know $e_T := (t_i, p_i)_{i \in I, t_i \in [-T,0]}$, the realization of the environment on the time interval $[-T,0]$, but we do not know the environment $\tilde{E}$ on the interval $(-\infty, -T)$ of the far away past: we only know that it is a realization of the a Poisson point process on $(-\infty, -T) \times (0, 1)$ with intensity measure $dt \times \mu$. Now, under the measure $\mathbb{P}^{e_T}$ the environment $e_T$ on $[-T,0]$ is fixed and we integrate with respect to the random environment $\tilde{E}$ on $(-\infty, -T)$. In addition, $\mathbb{P}$ simply denotes the annealed measure, i.e. when we integrate with respect to the environment $e$ on $(-\infty, 0)$. From Theorem 2.10 we have

$$\mathbb{E}[(1 - X(0))^n] = \mathbb{E}[(1 - X(\infty))^n] = w_n. \quad (7.1)$$

Our aim is to provide a formula for $\mathbb{E}^{e_T}[(1 - X(e_T,0))^n]$. Let $\tilde{e}$ denote a particular realization of the random environment $\tilde{E}$ on $(-\infty, -T)$. Let $\mathbb{E}^{\tilde{e},e_T}[(1 - X(\tilde{e},e_T,0))^n]$ denotes the quenched moment of order $n$ of $1 - X(\tilde{e},e_T,0)$, given the global environment $(\tilde{e},e_T)$ that is obtained by gluing together $\tilde{e}$ and $e_T$. Clearly, we have

$$\mathbb{E}^{e_T}[(1 - X(e_T,0))^n] = \int \mathbb{E}^{\tilde{e},e_T}[(1 - X(\tilde{e},e_T,0))^n] \times \mathbb{P}(\tilde{E} \in d\tilde{e}), \quad (7.2)$$

and

$$\mathbb{E}^{\tilde{e},e_T}[(1 - X(\tilde{e},e_T,0))^n] = \int \mathbb{E}^{e_T}[(1 - X(e_T,0))^n|X(e_T, -T) = x] \times \mathbb{P}(\tilde{e} \in X(\tilde{e}, -T) \in dx). \quad (7.3)$$

Let $-T < T_N < \cdots < T_1 < 0$ denote the jumping times of $e_T$ on $[-T, 0]$. Let the matrices $(A_n^{(e)}(e_T))_{1 \leq n \leq N}$ and the coefficients $C_{n,k}^{(e)}(e_T, T)$ be defined as in Theorem 2.14 from the environment $e_T$. By (2.15) in Theorem 2.14 we have

$$\mathbb{E}^{e_T}[(1 - X(e_T,0))^n|X(e_T, -T) = x] = \sum_{k=0}^{n-1} C_{n,k}^{(e)}(e_T, T) A_k(x).$$

Plugging this into (7.3) and using (2.12) we get
\[ E^{e,T}[(1 - X(e_T, 0))^n] = \sum_{k=0}^{n^2} C_{n,k}^{\uparrow} (e_T, T) E^e \left[ P_k^T (1 - X(\bar{e}, -T)) \right] \]
\[ = \sum_{j=0}^{n^2} \left( \sum_{k=0}^{n^2} C_{n,k}^{\uparrow} (e_T, T) u_{k,j} \right) E^e \left[ (1 - X(\bar{e}, -T))^j \right]. \]

Integrating both sides with respect to \( \bar{e} \) and using (7.2) we get

\[ E^{e,T}[(1 - X(e_T, 0))^n] = \sum_{j=0}^{n^2} \left( \sum_{k=0}^{n^2} C_{n,k}^{\uparrow} (e_T, T) u_{k,j} \right) E \left[ (1 - X(-T))^j \right]. \]

Here, \( E[(1 - X(-T))^j] \) is the annealed moment of order \( j \) of the proportion of individuals of type 1 at time \( -T \), when the environment is integrated on \((-\infty, -T)\). By translation invariance we have \( E[(1 - X(-T))^j] = E[(1 - X(0))^j] = w_j \), where the last equality comes from (7.1). Hence, the desired formula for the \( n \)-th moment of \( 1 - X(e_T, 0) \) is given by

\[ E^{e,T}[(1 - X(e_T, 0))^n] = \sum_{j=0}^{n^2} \left( \sum_{k=0}^{n^2} C_{n,k}^{\uparrow} (e_T, T) u_{k,j} \right) w_j. \]  \( (7.4) \)

**Appendix A. Some definitions and remarks on the Skorokhod topology**

For \( T > 0 \), we denote by \( \mathbb{D}_T \) the space of càdlàg functions in \([0, T]\) with values on \( \mathbb{R} \). Let \( C_{\uparrow}^T \) denote the set of increasing, continuous functions from \([0, T]\) onto itself. For \( \lambda \in C_{\uparrow}^T \), we set

\[ ||\lambda||^0_T := \sup_{0 \leq u < s \leq T} \left| \log \left( \frac{\lambda(s) - \lambda(u)}{s - u} \right) \right|. \]

We define the metric \( d_\uparrow^T \) in \( \mathbb{D}_T \) via

\[ d_\uparrow^T(f, g) := \inf_{\lambda \in C_{\uparrow}^T} \{ ||\lambda||^0_T \vee ||f - g \circ \lambda||_{T, \infty} \}, \]

where \( ||f||_{T, \infty} := \sup_{s \in [0, T]} |f(s)| \). The metric \( d_\uparrow^T \) induces the Skorokhod topology in \( \mathbb{D}_T \). An important feature is that the space \((\mathbb{D}_T^0, d_\uparrow^0)\) is separable and complete.

For \( T > 0 \), a function \( \omega \in \mathbb{D}_T \) is said to be pure-jump if for all \( t \in (0, T] \),

\[ \omega(t) - \omega(0) = \sum_{u \in (0, t]} \Delta \omega(u) < \infty, \]

where \( \Delta \omega(u) := \omega(u) - \omega(u^-), u \in [0, T] \). In the set of pure-jump functions it is sometimes convenient to work with the following metric

\[ d_\uparrow^T(\omega_1, \omega_2) := \inf_{\lambda \in C_{\uparrow}^T} \left\{ ||\lambda||^0_T \vee \sum_{u \in [0, T]} |\Delta \omega_1(u) - \Delta (\omega_2 \circ \lambda)(u)| \right\}. \]

The next result provides comparison inequalities between the metrics \( d_\uparrow^T \) and \( d_\uparrow^T \).

**Lemma A.1.** Let \( \omega_1 \) and \( \omega_2 \) be two pure-jump functions with \( \omega_1(0) = \omega_2(0) = 0 \), then

\[ d_\uparrow^T(\omega_1, \omega_2) \leq d_\uparrow^T(\omega_1, \omega_2). \]

If in addition, \( \omega_1 \) and \( \omega_2 \) are non-decreasing, and \( \omega_1 \) jumps exactly \( n \) times in \([0, T]\), then

\[ d_\uparrow^T(\omega_1, \omega_2) \leq (4n + 3)d_\uparrow^T(\omega_1, \omega_2). \]

**Proof.** Let \( \lambda \in C_{\uparrow}^T \) and set \( f := \omega_1 \) and \( g := \omega_2 \circ \lambda \). Since \( f \) and \( g \) are pure-jump functions with the same value at 0, we have, for any \( t \in [0, T] \),

\[ |f(t) - g(t)| = \left| \sum_{u \in [0, t]} (\Delta f(u) - \Delta g(u)) \right| \leq \sum_{u \in [0, t]} |\Delta f(u) - \Delta g(u)|. \]
The first inequality follows. Now, assume that ω1 and ω2 are non-decreasing and that ω1 has n jumps in
[0, T]. Let t1 < · · · < tn be the consecutive jumps of ω1. We first prove by induction that for any k ∈ [n]
\[\sum_{u \in [0, t_k]} |\Delta f(u) - \Delta g(u)| \leq (4k + 1)\|f - g\|_{t_k, \infty}.\]

Note that
\[\sum_{u \in [0, t_1]} |\Delta f(u) - \Delta g(u)| = \sum_{u \in [0, t_1]} \Delta g(u) + |\Delta f(t_1) - \Delta g(t_1)| \leq g(t_1) + 2\|f - g\|_{t_1, \infty} \leq 3\|f - g\|_{t_1, \infty},\]
which proves the result for k = 1. Now, assuming that the result is true for k ∈ [n − 1], we obtain
\[\sum_{u \in [0, t_k+1]} |\Delta f(u) - \Delta g(u)| = \sum_{u \in [0, t_k]} |\Delta f(u) - \Delta g(u)| + \sum_{u \in (t_k, t_{k+1})} \Delta g(u) + |\Delta f(t_{k+1}) - \Delta g(t_{k+1})|\]
\[\leq (4k + 1)\|f - g\|_{t_k, \infty} + g(t_{k+1}) - g(t_k) + 2\|f - g\|_{t_{k+1}, \infty}\]
\[= (4k + 1)\|f - g\|_{t_k, \infty} + g(t_{k+1}) - f(t_{k+1}) - (g(t_k) - f(t_k))\]
\[+ 2\|f - g\|_{t_{k+1}, \infty}\]
\[\leq (4k + 1)\|f - g\|_{t_k, \infty} + 4\|f - g\|_{t_{k+1}, \infty}\]
\[\leq (4(k + 1) + 1)\|f - g\|_{t_{k+1}, \infty}.\]

Hence, the result also holds for k + 1. This ends the proof of the claim. Finally, using the claim, we get
\[\sum_{u \in [0, T]} |\Delta f(u) - \Delta g(u)| = \sum_{u \in [0, t_n]} |\Delta f(u) - \Delta g(u)| + \sum_{u \in (t_n, T]} \Delta g(u)\]
\[\leq (4n + 1)\|f - g\|_{t_n, \infty} + g(T) - g(t_n)\]
\[= (4n + 1)\|f - g\|_{t_n, \infty} + (g(T) - f(T)) - (g(t_n) - f(t_n))\]
\[\leq (4n + 3)\|f - g\|_{T, \infty},\]
ending the proof. \(\square\)

**Appendix B. Bounded Lipschitz metric and weak convergence**

In what follows \((E, d)\) denotes a complete and separable metric space. It is well known that the topology of weak convergence of probability measures on \(E\) is induced by the Prokhorov metric. An alternative metric inducing this topology is given by the bounded Lipschitz metric. We recall its definition in this section.

**Definition B.1** (Lipschitz function). A real valued function \(F\) on \((E, d)\) is said to be Lipschitz if there is \(K > 0\) such that
\[|F(x) - F(y)| \leq Kd(x, y), \quad \text{for all } x, y \in E.\]

We denote by \(BL(E)\) the vector space of bounded Lipschitz functions on \(E\). The space \(BL(E)\) is equipped with the norm
\[\|F\|_{BL} := \sup_{x \in E} |F(x)| \vee \sup_{x, y \in E, x \neq y} \left\{\frac{|F(x) - F(y)|}{d(x, y)}\right\}, \quad F \in BL(E).\]

**Definition B.2** (Bounded Lipschitz metric). Let \(\mu, \nu\) be two probability measures on \(E\). The bounded Lipschitz distance between \(\mu\) and \(\nu\) is defined by
\[\varrho_E(\mu, \nu) := \sup \left\{\left|\int Fd\mu - \int Fd\nu\right| : F \in BL(E), \|F\|_{BL} \leq 1\right\}.
\]

It is not difficult to show that the bounded Lipschitz distance defines indeed a metric on the space of probability measures on \(E\). Moreover, the bounded Lipschitz distance metrizes the weak convergence of probability measures on \(E\), i.e.
\[\varrho_E(\mu_n, \mu) \xrightarrow{n \to \infty} 0 \iff \mu_n \xrightarrow{(d)_{n \to \infty}} \mu.\]
Acknowledgements

We are grateful to E. Baake for many interesting discussions. F. Cordero received financial support from Deutsche Forschungsgemeinschaft (CRC 1283 “Taming Uncertainty”, Project C1). Grégoire Véchambre acknowledges the support from Deutsche Forschungsgemeinschaft (CRC 1283 “Taming Uncertainty”, Project C1) for his visit to Bielefeld University in January 2018. Fernando Cordero acknowledges the support of NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai for his visit to NYU Shanghai in July 2018.

References

[1] E. Baake and A. Wakolbinger. Lines of descent under selection. *Journal of Statistical Physics*, 172(1):156–174, Jul. ISSN 1572-9613.

[2] E. Baake, U. Lenz, and A. Wakolbinger. The common ancestor type distribution of a Λ-Wright–Fisher process with selection and mutation. *Electron. Commun. Probab.*, 21:no. 59, 16 pp., 2016.

[3] E. Baake, F. Cordero, and S. Hummel. Lines of descent in the deterministic mutation-selection model with pairwise interaction. *arXiv e-prints*, 2018.

[4] B. Bah and E. Pardoux. The Λ-lookdown model with selection. *Stochastic Process. Appl.*, 125:1089–1126, 2015.

[5] V. Bansaye, T. G. Kurtz, and F. Simatos. Tightness for processes with fixed points of discontinuities and applications in varying environment. *Electron. Commun. Probab.*, 21:Paper No. 81, 9, 2016.

[6] V. Bansaye, M.-E. Caballero, and S. Méleard. Scaling limits of population and evolution processes in random environment. *Electron. J. Probab.*, 24:38 pp., 2019. doi: 10.1214/19-EJP262.

[7] M. Barczy, Z. Li, and G. Pap. Yamada–Watanabe results for stochastic differential equations with jumps. *Int. J. Stoch. Anal.*, pages Art. ID 460472, 23, 2015.

[8] N. Biswas, A. Etheridge, and A. Klimek. The spatial Lambda-Fleming–Viot process with fluctuating selection, 2018.

[9] J. J. Bull. Evolution of phenotypic variance. *Evolution*, 41(2):303–315, 1987. doi: 10.1111/j.1558-5646.1987.tb05799.x.

[10] R. Bürger and A. Gimelfarb. Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genetical Research*, 80(1):31–46, 2002. doi: 10.1017/S0016672302005682.

[11] J. Chetwynd-Diggle and A. Klimek. Rare mutations in the spatial Lambda-Fleming–Viot model in a fluctuating environment and super Brownian motion. 2019.

[12] F. Cordero. Common ancestor type distribution: a Moran model and its deterministic limit. *Stochastic Process. Appl.*, 127(2):590–621, 2017.

[13] F. Cordero, S. Hummel, and E. Schertzer. General selection models: Bernstein duality and minimal ancestral structures. *arXiv e-prints*, 2019.

[14] R. Durrett. *Probability Models for DNA Sequence Evolution*. Springer, New York, 2nd edition, 2008.

[15] A. M. Etheridge, R. C. Griffiths, and J. E. Taylor. A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theor. Popul. Biol.*, 78:77–92, 2010.

[16] C. Foucart. The impact of selection in the Λ-Wright–Fisher model. *Electron. Commun. Probab.*, 18:no. 72, 10 pp., 2013.

[17] Z. Fu and Z. Li. Stochastic equations of non-negative processes with jumps. *Stochastic Process. Appl.*, 120(3):306–330, 2010. doi: 10.1016/j.spa.2009.11.005.

[18] J. H. Gillespie. The effects of stochastic environments on allele frequencies in natural populations. *Theor. Popul. Biol.*, 3(3):241–248, 1972. ISSN 0040-5809. doi: https://doi.org/10.1016/0040-5809(72)90001-9.

[19] A. González Casanova and D. Spanò. Duality and fixation in Ξ-Wright–Fisher processes with frequency-dependent selection. *Ann. Appl. Probab.*, 28:250–284, 2018.

[20] A. González Casanova, D. Spanò, and M. Wilke Berenguer. The effective strength of selection in random environment. *arXiv e-prints*, 2019.

[21] P. Imkeller and N. Willrich. Solutions of martingale problems for Lévy-type operators with discontinuous coefficients and related SDEs. *Stochastic Process. Appl.*, 126:703–734, 2016.
[22] S. Jansen and N. Kurt. On the notion(s) of duality for Markov processes. *Probab. Surv.*, 11:59–120, 2014. doi: 10.1214/12-PS206.

[23] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. doi: 10.1007/978-1-4757-4015-8.

[24] S. Karlin and B. Benny Levikson. Temporal fluctuations in selection intensities: Case of small population size. *Theor. Popul. Biol.*, 6(3):383–412, 1974. doi: https://doi.org/10.1016/0040-5809(74)90017-3.

[25] S. Karlin and U. Liberman. Random temporal variation in selection intensities: One-locus two-allele model. *Journal of Mathematical Biology*, 2(1):1–17, Mar 1975. doi: 10.1007/BF00276012.

[26] S. Karlin and U. Lieberman. Random temporal variation in selection intensities: Case of large population size. *Theor. Popul. Biol.*, 6(3):355–382, 1974. ISSN 0040-5809. doi: https://doi.org/10.1016/0040-5809(74)90016-1.

[27] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Popul. Biol.*, 51(3):210–237, 1997.

[28] U. Lenz, S. Kluth, E. Baake, and A. Wakolbinger. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theor. Popul. Biol.*, 103:27–37, 2015.

[29] Z. Li and F. Pu. Strong solutions of jump-type stochastic equations. *Electron. Commun. Probab.*, 17:no. 33, 13, 2012. doi: 10.1214/ECP.v17-1915.

[30] C. Neuhauser. The ancestral graph and genealogy under frequency-dependent selection. *Theor. Popul. Biol.*, 56:203–214, 1999.

[31] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 145(2):519–534, February 1997.

[32] J. A. van Casteren. On martingales and Feller semigroups. *Results in Mathematics*, 21:274–288, 1992.