MORAN MODELS AND WRIGHT–FISHER DIFFUSIONS WITH SELECTION AND MUTATION IN A ONE-SIDED RANDOM ENVIRONMENT

FERNANDO CORDERO ![ORCID],* *Bielefeld University*

GRÉGOIRE VÉCHAMBRE ![ORCID],** *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

Abstract

Consider a two-type Moran population of size N with selection and mutation, where the selective advantage of the fit individuals is amplified at extreme environmental conditions. Assume selection and mutation are weak with respect to N, and extreme environmental conditions rarely occur. We show that, as $N \to \infty$, the type frequency process with time sped up by N converges to the solution to a Wright–Fisher-type SDE with a jump term modeling the effect of the environment. We use an extension of the *ancestral selection graph* (ASG) to describe the genealogical picture of the model. Next, we show that the type frequency process and the line-counting process of a pruned version of the ASG satisfy a moment duality. This relation yields a characterization of the asymptotic type distribution. We characterize the ancestral type distribution using an alternative pruning of the ASG. Most of our results are stated in annealed and quenched form.

Keywords: Wright–Fisher diffusion; Moran model; random environment; ancestral selection graph; duality

2020 Mathematics Subject Classification: Primary 82C22; 92D15
Secondary 60J25; 60J27

1. Introduction

The *Wright–Fisher diffusion with mutation and selection* describes the evolution of the type composition of an infinite two-type haploid population, which is subject to mutation and selection. Fit individuals reproduce at rate $1 + \sigma$, $\sigma \geq 0$, whereas unfit ones reproduce at rate 1. In addition, individuals mutate at rate θ to the fit type with probability $\nu_0 \in [0, 1]$, and to the unfit type with probability $\nu_1 := 1 - \nu_0$. The proportion of fit individuals evolves forward in time according to the stochastic differential equation (SDE)

$$\mathrm{d}X(t) = (\theta \nu_0(1 - X(t)) - \theta \nu_1 X(t) + \sigma X(t)(1 - X(t))) \, \mathrm{d}t$$
$$+ \sqrt{2X(t)(1 - X(t))} \, \mathrm{d}B(t), \quad t \geq 0, \qquad (1.1)$$

where $(B(t))_{t \geq 0}$ is a standard Brownian motion. The solution to (1.1) arises as the *diffusion approximation* of (properly normalized) continuous-time Moran models and discrete-time Wright–Fisher models. In the neutral case, it also appears as the limit of a large class of Cannings models (see [33]). The genealogical counterpart to X is the *ancestral selection graph* (ASG), which is a branching–coalescing process coding the potential ancestors of an untyped sample of the population at present. It was introduced by Krone and Neuhauser in [29, 35] and later extended to models evolving under general neutral reproduction mechanisms (see [2, 15]), and to general forms of frequency-dependent selection (see [1, 12, 19, 34]).

For $\theta = 0$, the process $(R(t))_{t \geq 0}$ that counts the lines in the ASG is moment dual with $1 - X$, i.e. for every $n \in \mathbb{N}$, $x \in [0, 1]$, and $t \geq 0$ we have

$$\mathbb{E}\left[(1 - X(t))^n \mid X(0) = x\right] = \mathbb{E}\left[(1 - x)^{R(t)} \mid R(0) = n\right]. \qquad (1.2)$$

This relation yields an expression for the absorption probability of X at 0 in terms of the stationary distribution of R. For $\theta > 0$, there are two variants of the ASG that dynamically resolve mutation events and encode relevant information about the model: *the killed ASG* and *the pruned lookdown ASG*. The killed ASG was introduced in [3] to determine whether a sample consists only of unfit individuals. Its line-counting process extends Equation (1.2) to the case $\theta > 0$ [3, Proposition 1] (see [13, Proposition 2.2] for a generalization). This allows one to characterize the stationary distribution of X. The pruned lookdown ASG in turn was introduced in [31] (see [2, 11] for extensions) as a tool for determining the *ancestral type distribution*, i.e. the type distribution of the individuals that have been successful in the long run.

In many biological situations the strength of selection fluctuates in time. 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. [8, 9, 18, 25–27, 36]), and it is currently experiencing renewed interest (see e.g. [4, 7, 10, 20–22]). In this paper, we consider the 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.). As an example, consider a population consisting of fit and unfit individuals which is subject to catastrophes. Assume that only fit individuals are resistant to the catastrophes. Hence, shortly after a catastrophe the population may drop below its carrying capacity and subsequently grow quickly. Since fit individuals have a reproductive advantage, it is likely that their relative frequency will grow fast after a catastrophe. One may also think of a population consisting of individuals that are specialized to high temperatures, as well as wild-type individuals accustomed but not specialized to them. The environment is characterized by moderately high temperatures, present most of the time, and short periods of extreme heat. It is then likely that specialized individuals have a (slight) reproductive advantage under moderate temperatures, and a more prominent advantage at extreme temperatures.

To model the previously described scenario we use a two-type Moran population with mutation and selection, immersed in a varying environment. The environment is modeled via a countable collection of points $\zeta := (t_i, p_i)_{i \in I}$ in $(-\infty, \infty) \times (0, 1)$, satisfying that $\sum_{t_i \in [s,t]} p_i < \infty$ for all $s < t$. Each t_i represents the time of an instantaneous environmental change; the *peak* p_i models the strength of this event: at time t_i each fit individual independently reproduces with probability p_i. Each offspring replaces a different individual in the population, so that the population size remains constant. The summability of the peaks ensures that the number of reproductions in any compact time interval is almost surely finite. In this context, we show that the type frequency process is continuous with respect to the environment. The proof uses coupling techniques that uncover the effect of small environmental changes.

Next, we consider a random environment given by a Poisson point process on $(-\infty, \infty) \times (0, 1)$ with intensity measure $dt \times \mu$, where dt stands for the Lebesgue measure and μ is a measure on $(0, 1)$ satisfying $\int x\mu(dx) < \infty$. Then we let population size grow to infinity, and we show that, in an appropriate parameter regime, the fit-type frequency process converges to the solution to the SDE

$$dX(t) = (\theta\nu_0(1 - X(t)) - \theta\nu_1 X(t))\, dt + X(t-)(1 - X(t-))dS(t)$$
$$+ \sqrt{2X(t)(1 - X(t))}dB(t), \quad t \geq 0, \qquad (1.3)$$

where $S(t) := \sigma t + J(t)$, and J is a pure-jump subordinator with Lévy measure μ, independent of B, which represents the cumulative effect of the environment. We refer to X as the *Wright–Fisher diffusion in random environment*. We prove the convergence in an annealed setting, i.e. when the environment is random. For environments given by compound Poisson processes, we show that the convergence also holds in a quenched sense, i.e. when a realization of the environment is fixed. For $\theta = 0$, Equation (1.3) is a particular case of [4, Equation 3.3], which arises as the large-population limit of a family of discrete-time Wright–Fisher models [4, Theorem 3.2].

Next, we generalize the construction of the ASG, the killed ASG, and the pruned lookdown ASG to incorporate the effect of the environment. In the annealed case, we establish a relation between X, the line-counting process of the killed ASG, and the total increment of the environment; we refer to this relation as a *reinforced moment duality*. The latter is a central tool for characterizing the asymptotic type frequencies. We also express the ancestral type distribution in terms of the line-counting process of the pruned lookdown ASG. Analogous results are obtained in the more involved quenched setting.

As an application of our results, we compare the long-term behavior of two Wright–Fisher diffusions without mutations, the first one having parameter $\sigma = 0$ and an environment with Lévy measure μ, and the second one having parameter $\sigma = \int_{(0,1)} y\mu(dy)$ and no environment. We prove that the probability of fixation of the fit type is smaller under the first model than under the second one, provided that the initial frequency of fit individuals is sufficiently large; see Proposition 2.2.

The analysis of a more realistic scenario where environmental changes are not always favorable to the same type cannot be done via the methods presented in this paper. The main reason is that in such a setting the frequency process does not admit a moment dual. To circumvent this problem one has to take into account the whole combinatorics of the ASG, which is a cumbersome object. This is the object of a forthcoming study.

We would also like to mention the parallel development by González Casanova *et al.* in [20]. They study the accessibility of the boundaries and the fixation probabilities of a generalization of the SDE (1.3) with $\theta = 0$. The paper [20] makes use only of the ASG and does not cover the case $\theta > 0$, where the killed and the pruned lookdown ASG play a pivotal role. Moreover, the reinforced moment duality and all the results obtained in the quenched setting are to the best of our knowledge new.

This article is organized as follows. An outline of the paper containing our main results is given in Section 2. In Section 3 we prove the continuity of the type frequency process in the Moran model with respect to the environment, that (1.3) is well-posed, and that it arises in the large-population limit of the type frequency process of a sequence of Moran models. In Section 4 we give more detailed definitions of the ASG, the killed ASG, and the pruned lookdown ASG. Section 5 is devoted to the proofs of (i) the annealed moment duality between the process X and the line-counting process of the killed ASG, (ii) the long-term behavior of

the annealed type frequency process, and (iii) the annealed ancestral type distribution. The quenched versions of these results are proved in Section 6. Section 7 provides additional (quenched) results for environments having finitely many jumps in any compact time interval.

2. Description of the model and main results

Notation. The positive integers are denoted by $\mathbb{N}$, and we set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For $m \in \mathbb{N}$,

$$[m] := \{1, \dots, m\}, \quad [m]_0 := [m] \cup \{0\}, \quad \text{and} \quad]m] := [m] \setminus \{1\}.$$

For $x, y \in \mathbb{R}$, we define $x \wedge y := \min\{x, y\}$, $x \vee y := \max\{x, y\}$, and $(x)_+ := x \vee 0$. For $s < t$, we denote by $\mathbb{D}_{s,t}$ (resp. $\mathbb{D}$) the space of càdlàg functions from $[s, t]$ (resp. $\mathbb{R}$) to $\mathbb{R}$, which is endowed with the Billingsley metric inducing the J_1-Skorokhod topology and makes the space complete (see Appendix A.1). For any Borel set $S \subset \mathbb{R}$, denote by $\mathcal{M}_f(S)$ (resp. $\mathcal{M}_1(S)$) the set of finite (resp. probability) measures on S. We use $\xrightarrow{(d)}$ to denote convergence in distribution of random variables and $\xRightarrow{(d)}$ for convergence in distribution of càdlàg processes.

For $n \in \mathbb{N}_0$ and $k, m \in [n]_0$, we write $K \sim \mathrm{Hyp}(n, m, k)$ if K is a hypergeometric random variable with parameters n, m, and k, i.e.

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

For $x \in [0, 1]$ and $n \in \mathbb{N}$, we write $B \sim \mathrm{Bin}(n, x)$ if B is a binomial random variable with parameters n and x, i.e.

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

Relevant notation introduced in the next sections is collected in Appendix B.

2.1. Moran models in deterministic pure-jump environments

Consider a population of size N with two types, 0 and 1, subject to mutation and selection influenced by a deterministic environment. The latter is modeled by an at most countable collection $\zeta := (t_k, p_k)_{k \in I}$ of points in $(-\infty, \infty) \times (0, 1)$ satisfying for any $s, t \in \mathbb{R}$ with $s \leq t$ that

$$\sum_{t_k \in [s,t]} p_k < \infty. \tag{2.1}$$

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 independently mutates at rate $\theta_N \geq 0$ with probability $\nu_0 \in [0, 1]$ (resp. $\nu_1 := 1 - \nu_0$) to type 0 (resp. 1). Reproduction occurs independently from mutation. Individuals of type 1 reproduce at rate 1, whereas individuals of type 0 reproduce at rate $1 + \sigma_N$, $\sigma_N \geq 0$. (The subscript N in the parameters σ_N and θ_N emphasizes their dependence on N. In Theorem 2.2 we will require that they are asymptotically proportional to $1/N$.) Thus, we refer to type 0 (resp. type 1) as the *fit* (resp. *unfit*) type. In addition, at time t_k each type-0 individual independently reproduces with probability p_k. At reproduction times, (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 population present before the reproduction event (including the parents) to die, keeping the size of the population constant.
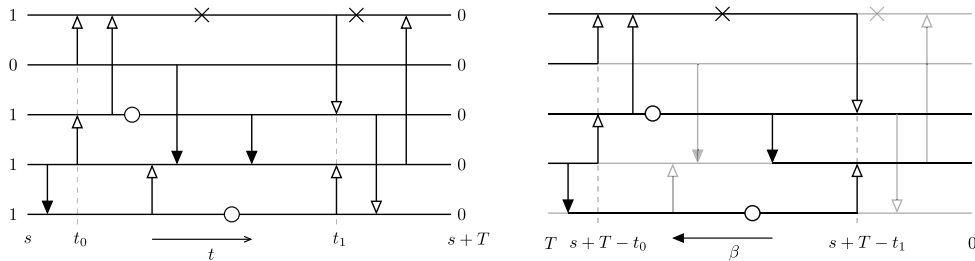
FIGURE 1. Left: a realization of the Moran IPS; time runs forward from left to right; the environment has peaks at times t_0 and t_1. Right: the ASG that arises from the second and third lines (from bottom to top) in the left picture, with the potential ancestors drawn in black; time runs backward from right to left; backward time $\beta \in [0, T]$ corresponds to forward time $t = s + T - \beta$.

Graphical representation. In the absence of environmental factors (i.e. $\zeta = \emptyset$), it is classical to describe the evolution of the population by means of the graphical representation as an interacting particle system (IPS). This decouples the randomness of the model due to the initial type configuration from the randomness due to mutations and reproductions. We now extend the graphical representation to incorporate the effect of the environment (see Section 3.1 for a more detailed description).

In the graphical representation, individuals are represented by horizontal lines at levels $i \in [N]$ (see Figure 1). Time runs forward from left to right. Potential reproduction events are depicted by arrows, with the (potential) parent at the tail and the offspring at the tip. We distinguish between neutral and selective arrows. *Neutral arrows* have a filled arrowhead; they occur at rate $1/N$ per pair of lines. *Selective arrows* have an open arrowhead; they occur in two independent ways: first, they occur at rate σ_N/N per pair of lines, and second, at any time t_k, $k \in I$, a random number $n_k \sim \text{Bin}(N, p_k)$ of lines shoot selective arrows to n_k individuals in the population. Furthermore, beneficial (deleterious) mutations, depicted as circles (crosses), occur at rate $\theta_N \nu_0$ (at rate $\theta_N \nu_1$) per line.

Note that for any $s < t$, the number of non-environmental graphical elements present in $[s, t]$ is almost surely finite. Moreover, thanks to Assumption (2.1), we have

$$\mathbb{E}\left[\sum_{t_k \in [s,t]} n_k\right] = N \sum_{t_k \in [s,t]} p_k < \infty. \tag{2.2}$$

Hence, $\sum_{t_k \in [s,t]} n_k < \infty$ almost surely, i.e. the number of arrows in $[s, t]$ due to peaks of the environment is almost surely finite.

Once the graphical elements in $[s, t] \times [N]$ are drawn, we specify the initial conditions by assigning types to the N lines at time s and propagate them forward in time according to the following rules: the type of a line right after a circle (resp. cross) is 0 (resp. 1). In particular, if the type before the circle (resp. cross) is 0 (resp. 1), the mutation is silent. Type 0 propagates through neutral arrows *and* selective arrows; type 1 propagates only through neutral arrows.

Reading off ancestries in the Moran model. The *ancestral selection graph* (ASG) was introduced by Krone and Neuhauser in [29] (see also [35]) to study the genealogical relations in the diffusion limit of the Moran model with mutation and selection. In what follows, we briefly explain how to adapt this construction to the Moran model in deterministic environment.

Consider a realization of the IPS associated to the Moran model in the environment $\zeta := (t_k, p_k)_{k \in I}$ in the time interval $[s, s + T]$. Fix an untyped sample of n individuals at time
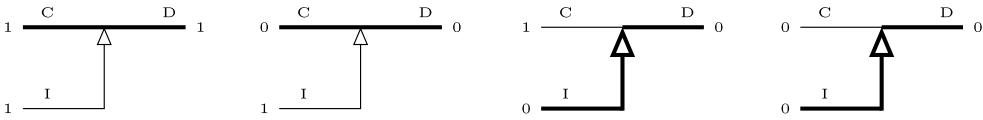
FIGURE 2. The descendant line (D) splits into the continuing line (C) and the incoming line (I). The incoming line is ancestral if and only if it is of type 0. The true ancestral line is drawn in bold.

$s + T$ and trace backward in time (from right to left in Figure 1) the lines of their potential ancestors (i.e. the lines that are ancestral to the sample for some type-configuration at time s), ignoring the effect of mutations; the backward time $\beta \in [0, T]$ corresponds to the forward time $t = s + T - \beta$. We do this as follows. When a neutral arrow joins two individuals in the current set of potential ancestors, the two lines coalesce into a single one at the tail of the arrow. When a neutral arrow hits a potential ancestor from outside the current set of potential ancestors, 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 peak times, multiple lines in the ASG can be hit by selective arrows, and therefore, multiple branching and collision events can occur simultaneously. Mutations are superposed on the lines of the ASG.

The object arising under this procedure up to time $\beta = T$ is called the *Moran-ASG in* $[s, s + T]$ *under the environment* ζ. It contains all the lines that are potentially ancestral (ignoring mutation events) to the lines sampled at time $t = s + T$; see Figure 1. Note that, since the number of events occurring in $[s, s + T]$ is almost surely finite (see (2.2)), the Moran-ASG in $[s, s + T]$ is well-defined.

Given an assignment of types to the lines present in the ASG at time $t = s$, we can extract the true genealogy and determine the types of the sampled individuals at time $t = s + T$. To this end, 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 2. This rule is called the *pecking order*. Proceeding in this way, the types in $[s, s + T]$ are determined along with the true genealogy.

Evolution of the type composition. Consider the set $\mathbb{D}^\star$ of non-decreasing functions $\omega \in \mathbb{D}$ satisfying the following:

(i) for all $s < t \in \mathbb{R}$, $\Delta\omega(t) := \omega(t) - \omega(t-) \in [0, 1)$ and $\sum_{u \in [s,t]} \Delta\omega(u) < \infty$;

(ii) ω is pure-jump, i.e. for all $s < t$, $\omega(t) = \omega(s) + \sum_{u \in (s,t]} \Delta\omega(u)$.

Note that the set of environments $\zeta := (t_k, p_k)_{k \in I}$ satisfying (2.1) can be identified with the set of functions $\omega \in \mathbb{D}^\star$ with $\omega(0) = 0$. Indeed, for any $\omega \in \mathbb{D}^\star$, the collection of points $\{(t, \Delta\omega(t)) : \Delta\omega(t) > 0)\}$ is countable and satisfies (2.1) (the same collection is obtained if we add a constant to ω; this is why we set $\omega(0) = 0$). Conversely, for any $\zeta := (t_k, p_k)_{k \in I}$, the function $\omega : \mathbb{R} \to \mathbb{R}$ defined via

$$\omega(t) := \sum_{t_k \in (0,t]} p_k \quad \text{for } t \geq 0, \quad \text{and} \quad \omega(t) := -\sum_{t_k \in (t,0]} p_k \quad \text{for } t < 0,$$

belongs to $\mathbb{D}^\star$. For this reason, we often abuse notation and refer to the elements of $\mathbb{D}^\star$ as environments. In addition, an environment $\omega \in \mathbb{D}^\star$ is said to be *simple* if ω has only a finite number of jumps in any compact time interval. We denote by **0** the environment corresponding to $\zeta = \emptyset$ and refer to it as the *null environment*.

We denote by $Z_N^\omega(t)$ the number of fit individuals at time t in a Moran population of size N subject to the environment $\omega \in \mathbb{D}^\star$. We refer to $Z_N^\omega := (Z_N^\omega(t))_{t \geq 0}$ as *the quenched fit-counting process*. In particular, $Z_N^{\mathbf{0}}$ is just the continuous-time Markov chain on $[N]_0$ with infinitesimal generator

$$
\mathcal{A}_N^0 f(n) := \left[(1 + \sigma_N) \frac{n(N-n)}{N} + \theta_N \nu_0 (N-n) \right] (f(n+1) - f(n))
$$
$$
+ \left[\frac{n(N-n)}{N} + \theta_N \nu_1 n \right] (f(n-1) - f(n)).
$$

Note that if ω has a jump at time t, the number $n(t)$ of individuals placing offspring is a binomial random variable with parameters $Z_N^\omega(t-)$ and $\Delta\omega(t)$. Since the $n(t)$ individuals that will be replaced are chosen uniformly at random, the additional number of fit individuals after the reproduction event is a hypergeometric random variable with parameters N, $N - Z_N^\omega(t-)$, and $n(t)$. Therefore, the dynamic of Z_N^ω is as follows. Recall that in any finite time interval, the number of environmental reproductions is almost surely finite. Thus, we can define $(S_i)_{i \in \mathbb{N}}$ as the increasing sequence of times at which environmental reproductions take place. We set $S_0 := 0$. By construction, $(S_i)_{i \in \mathbb{N}}$ is Markovian and its transition probabilities are given by

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

If $Z_N^\omega(0) = n \in [N]_0$, then Z_N^ω evolves in $[0, S_1)$ as $Z_N^{\mathbf{0}}$ started at n. For $i \in \mathbb{N}$, if $Z_N^\omega(S_i -) = k$, then $Z_N^\omega(S_i) = k + H(N, N-k, \tilde{B}_i(k))$, where the random variables $H(N, N-k, b) \sim \mathrm{Hyp}(N, N-k, b)$, $b \in [k]_0$, and $\tilde{B}_i(k)$ are independent, and $\tilde{B}_i(k)$ is a binomial random variable with parameters k and $\Delta\omega(S_i)$ conditioned to be positive. Then, Z_N^ω evolves in $[S_i, S_{i+1})$ as $Z_N^{\mathbf{0}}$ started at $Z_N^\omega(S_i)$.

Let us fix $T > 0$. We end this section with our first main result, which provides the continuity in $[0, T]$ of the *fit-counting* process with respect to the environment. Note that the restriction of the environment to $[0, T]$ can be identified with an element of

$$
\mathbb{D}_T^\star := \big\{ \omega \in \mathbb{D}_{0,T} : \omega(0) = 0, \ \Delta\omega(t) \in [0, 1) \text{ for all } t \in [0, T], \ \omega \text{ is }
$$
$$
\text{non-decreasing and pure-jump} \big\}. \tag{2.3}
$$

Moreover, we equip $\mathbb{D}_T^\star$ with the metric $d_T^\star$ defined in Appendix A.1, Equation (A.3).

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

$$
(Z_N^{\omega_k}(t))_{t \in [0,T]} \xrightarrow[k \to \infty]{(d)} (Z_N^\omega(t))_{t \in [0,T]}.
$$

Theorem 2.1 is proved in Section 3.1.

2.2. Moran models in subordinator-driven environments

In contrast to Section 2.1, we consider here a random environment given by a Poisson point process $(t_i, p_i)_{i \in I}$ on $(-\infty, \infty) \times (0, 1)$ with intensity measure $dt \times \mu$, where dt stands for the Lebesgue measure and μ is a measure on $(0, 1)$ satisfying

$$\int_{(0,1)} x \mu(dx) < \infty. \tag{2.4}$$

The latter implies that $(t_i, p_i)_{i \in I}$ almost surely satisfies Assumption (2.1). In particular, setting $J(t) := \sum_{t_i \in (0,t]} p_i$ for $t \geq 0$ and $J(t) := -\sum_{t_i \in (t,0]} p_i$ for $t < 0$, we have $J \in \mathbb{D}^\star$ almost surely. Moreover, by the Lévy–Itô decomposition, $(J(t))_{t \in \mathbb{R}}$ is a pure-jump subordinator with Lévy measure μ. If the measure μ is finite, then J is a compound Poisson process, and thus the environment J is almost surely simple.

We will see in Section 3.1 that, using the graphical representation, one can simultaneously construct Moran models for any $\omega \in \mathbb{D}^\star$. Now, consider an independent pure-jump subordinator $(J(t))_{t \geq 0}$, with Lévy measure μ on $(0, 1)$ satisfying (2.4). Thanks to Theorem 2.1 the process $Z_N^J := (Z_N^J(t))_{t \geq 0}$ is well-defined. We refer to Z_N^J as the *annealed fit-counting process*. By definition, we have

$$\mathbb{P}(Z_N^J \in \cdot) = \int \mathbb{P}(Z_N^\omega \in \cdot) \mathbb{P}(J \in d\omega).$$

In other words, $\mathbb{P}(Z_N^\omega \in \cdot)$ is the law of Z_N^J conditionally on a realization ω of the environment (i.e. $\mathbb{P}(Z_N^\omega \in \cdot) = \mathbb{P}(Z_N^J \in \cdot \mid J = \omega)$) and is classically referred to as the *quenched measure*, while $\mathbb{P}(Z_N^J \in \cdot)$ integrates the effect of the random environment and is classically referred to as the *annealed measure*.

The process Z_N^J is a continuous-time Markov chain on $[N]_0$ with generator

$$\mathcal{A}_N f(n) := \mathcal{A}_N^0 f(n) + \int_{(0,1)} \left(\mathbb{E}\left[f(n + \mathcal{H}(N, N - n, B_n(u))) \right] - f(n) \right) \mu(du), \quad n \in [N_0],$$

where $B_n(u) \sim \text{Bin}(n, u)$, and for any $i \in [n]_0$, $\mathcal{H}(N, N - n, i) \sim \text{Hyp}(N, N - n, i)$ are independent.

The dynamic of the graphical representation is as follows. For each $i, j \in [N]$ with $i \neq j$, selective (resp. neutral) arrows from level i to level j appear at rate σ_N/N (resp. $1/N$). For each $i \in [N]$, open circles (resp. crosses) appear at level i at rate $\theta_N \nu_0$ (resp. $\theta_N \nu_1$). For each $k \in [N]$, every group of k lines is subject to simultaneous potential reproductions at rate

$$\sigma_{N,k} := \int_{(0,1)} y^k (1-y)^{N-k} \mu(dy),$$

resulting in the appearance of k selective arrows from the lines of this group (of potential parents) to the lines of a group of size k (the potential descendants) that is chosen uniformly at random among subsets of size k of the N individuals. The k selective arrows are drawn uniformly at random from the k potential parents to the k potential descendants. Recall that only type 0 propagates through selective arrows, while both types propagate through neutral arrows. The appearance of a selective arrow is therefore silent when the potential parent at its tail is of type 1.
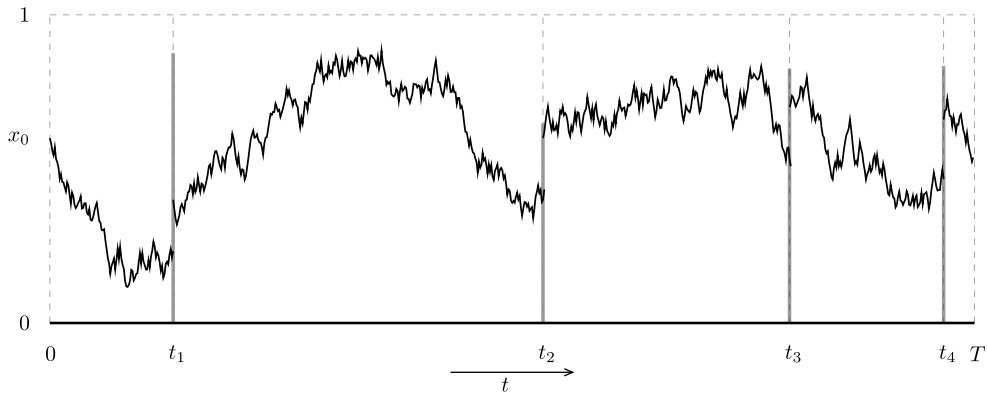
FIGURE 3. An illustration of a path of the process X^ω in the interval $[0, T]$. The grey vertical lines represent the peaks of the environment ω; t_1, t_2, t_3, and t_4 are the jump times of ω.

2.3. The Wright–Fisher diffusion in random environment

In this section we are interested in the Wright–Fisher diffusion in random environment described in the introduction as the solution to the SDE (1.3). In Section 3 we will see that, indeed, for any $x_0 \in [0, 1]$, this SDE has a pathwise unique strong solution starting at x_0 (see Proposition 3.3).

Consider a pure-jump subordinator $J = (J(s))_{s \geq 0}$ with Lévy measure satisfying (2.4) and an independent standard Brownian motion $B = (B(s))_{s \geq 0}$. For any $T > 0$, the solution to (1.3) in $[0, T]$ is a measurable function of $(B(s), J(s))_{s \in [0, T]}$, which we denote by $F(B, J)$. A regular version of the conditional law $\mathbb{P}(F(B, J) \in \cdot \mid J = \omega)$ of $F(B, J)$ given J is classically referred to as the *quenched probability measure*. It is defined for almost every realization ω of J. $\mathbb{P}(F(B, J) \in \cdot)$ integrates the effect of the random environment and is classically referred to as the *annealed measure*. As before, the quenched and annealed measures are related via

$$\mathbb{P}(F(B, J) \in \cdot) = \int \mathbb{P}(F(B, J) \in \cdot \mid J = \omega)\, \mathbb{P}(J \in \mathrm{d}\omega).$$

We write X and X^ω for the solutions to (1.3) under the annealed and quenched measures, respectively. For ω simple, the process X^ω starting at x_0 can be alternatively defined as follows. Denote by $t_1 < \cdots < t_k$ the consecutive jump times of ω in $[0, T]$, and set $t_0 := 0$ and $X^\omega(0) := x_0$. In the intervals $[t_i, t_{i+1})$, X^ω evolves as the solution to (1.1) starting at $X^\omega(t_i)$. Moreover, if $X^\omega(t_i -) = x$, then $X^\omega(t_i) := x + x(1 - x)\Delta\omega(t_i)$; see Figure 3 for an illustration.

The next result states the convergence of the type frequency process in the Moran model to the Wright–Fisher diffusion in random environment, as population size grows to ∞ and time is suitably accelerated.

Theorem 2.2. (Convergence.) *Assume that $N\sigma_N \to \sigma$ and $N\theta_N \to \theta$ as $N \to \infty$, for some $\sigma, \theta \geq 0$ (weak selection, weak mutation).*

1. *Let J be a pure-jump subordinator with Lévy measure μ in $(0,1)$, and set $J_N(t) := J(t/N)$, $t \geq 0$. Define the process $(X_N(t))_{t \geq 0}$ via $X_N(t) := Z_N^{J_N}(Nt)/N$, $t \geq 0$. If $X_N(0) \xrightarrow[N \to \infty]{(d)} x_0$, then*

$$(X_N(t))_{t\geq 0} \xrightarrow[N\to\infty]{(d)} (X(t))_{t\geq 0},$$

where X is the unique pathwise solution to (1.3) with $X(0) = x_0$.

2. Let $\omega \in \mathbb{D}^\star$ be a simple environment and set $\omega_N(t) := \omega(t/N)$, $t \geq 0$. Define the process $(X_N^\omega(t))_{t\geq 0}$ via $X_N^\omega(t) := Z_N^{\omega_N}(Nt)/N$, $t \geq 0$. If $X_N^\omega(0) \xrightarrow[N\to\infty]{(d)} x_0$, then

$$(X_N^\omega(t))_{t\geq 0} \xrightarrow[N\to\infty]{(d)} (X^\omega(t))_{t\geq 0},$$

with X^ω starting at x_0.

The proof of Theorem 2.2 is given in Section 3.2. The reason for using the environment J_N or ω_N is to compensate for the fact that time is sped up by a factor of N. In this way, X_N and X share the same environment.

Remark 2.1. The result analogous to Theorem 2.2(1) in the context of discrete-time Wright–Fisher models without mutations is covered by the fairly general result [4, Theorem 3.2] (see also [20, Theorem 2.12]).

Remark 2.2. If J is a compound Poisson process, then almost every environment is simple. In this case, according to Theorem 2.2(2), the quenched convergence holds for almost every environment (with respect to the law of J). We conjecture that this is true for general J. In Proposition 3.4 we show that the sequence $(X_N^\omega)_{N\geq 1}$ is tight for any environment ω. Hence, it would suffice to prove the continuity of $\omega \mapsto X^\omega$ to obtain the desired convergence. Unfortunately, since the diffusion term in (1.3) is not Lipschitz, the standard techniques used to prove this type of result fail. Developing new techniques to cover non-Lipschitz diffusion coefficients is beyond the scope of this paper.

2.4. The ancestral selection graph in random/deterministic environment

The aim of this section is to associate an ASG to the Wright–Fisher diffusion in random/deterministic environment. In contrast to the Moran model setting described in Section 2.1, it is not straightforward to set up a graphical representation for the forward process. To circumvent this problem, we proceed as follows. We first consider the graphical representation of a Moran model with parameters σ/N, θ/N, ν_0, ν_1, and environment $\omega_N(\cdot) = \omega(\cdot/N)$, and we speed up time by N. Next, we sample n individuals at time T and we construct the ASG as in Section 2.1.

Now, replace ω by a pure-jump subordinator J with Lévy measure μ supported in $(0,1)$. Note that the Moran-ASG in $[0, T]$ evolves according to the time reversal of J. The latter is the subordinator $\bar{J}^T := (\bar{J}^T(\beta))_{\beta\in[0,T]}$ with $\bar{J}^T(\beta) := J(T) - J((T-\beta)-)$, which has the same law as J (its law does not depend on T). A simple asymptotic analysis of the rates and probabilities for the possible events leads to the following definition.

Definition 2.1. *(The annealed/quenched ASG.)* The *annealed ancestral selection graph* $\mathcal{G} := (\mathcal{G}(\beta))_{\beta\geq 0}$ with parameters σ, θ, ν_0, ν_1, and environment driven by a pure-jump subordinator with Lévy measure μ, associated to a sample of size n of the population at time T is the branching–coalescing particle system starting with n lines and with the following dynamic:

  (i) Each line independently splits at rate σ into an incoming line and a continuing line.

  (ii) Every given pair of lines independently coalesces into a single one at rate 2.
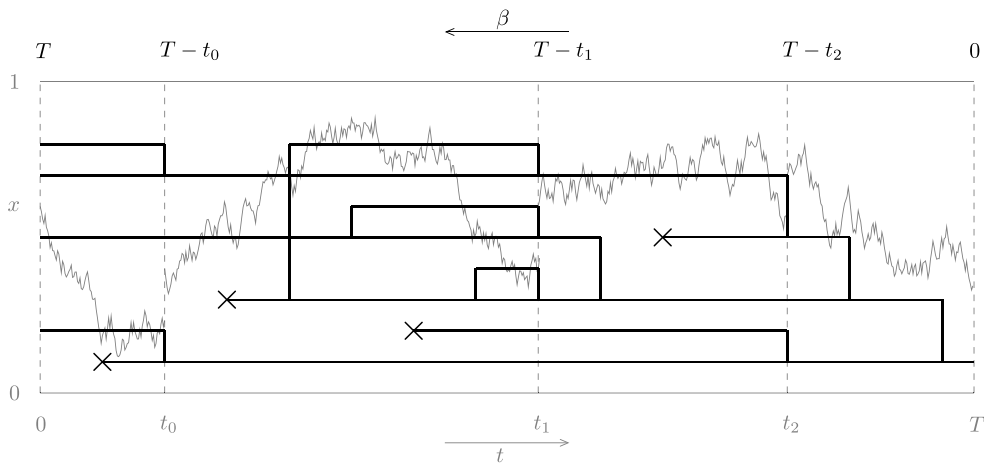
FIGURE 4. Illustration of a path of the process X^ω (grey path) and the killed ASG $\mathcal{G}^\omega_T$ (black lines) embedded in the same picture. Forward time t runs from left to right; backward time $\beta := T - t$ runs from right to left. The environment ω jumps at forward times t_0, t_1, and t_2.

(iii) If m is the current number of lines in the ASG, every group of k lines independently experiences a *simultaneous branching* at rate

$$\sigma_{m,k} := \int_{(0,1)} y^k (1 - y)^{m-k} \mu(\mathrm{d}y), \tag{2.5}$$

i.e. each line in the group splits into an incoming line and a continuing line.

(iii) Each line is independently decorated by a beneficial mutation at rate $\theta \nu_0$.

(v) Each line is independently decorated by a deleterious mutation at rate $\theta \nu_1$.

Let $\omega : \mathbb{R} \to \mathbb{R}$ be a fixed environment. The *quenched ancestral selection graph* with parameters σ, θ, ν_0, ν_1, and environment ω of a sample of size n at time T is a branching–coalescing particle system $\mathcal{G}^\omega_T := (\mathcal{G}^\omega_T(\beta))_{\beta \geq 0}$ starting at $\beta = 0-$ with n lines and evolving as the annealed ASG but with (iii) replaced by the following:

(iii′) If at time β we have $\Delta\omega(T - \beta) > 0$, then any line splits into two lines, an incoming line and a continuing line, with probability $\Delta\omega(T - \beta)$, independently from the other lines.

See Figure 4 for an illustration of the type frequency process X^ω and the killed ASG $\mathcal{G}^\omega_T$. The branching–coalescing system $\mathcal{G}^\omega_T$ is clearly well-defined for ω simple. The justification of the previous definition for general environments is more involved and will be given in Section 4.1.

Remark 2.3. In the Moran model, a neutral arrow appears from line A to line B at rate $1/N$ and from line B to line A at the same rate. Two lines are thus connected by a neutral arrow at rate $2/N$, which explains the rate of coalescence events in (ii).

2.5. Type frequency via the killed ASG

The aim of this section is to relate the type-0 frequency process X to the ASG. To this end, assume that the proportion of fit individuals at time 0 is equal to $x \in [0, 1]$. Conditionally

on $X(T)$, the probability of independently sampling 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 randomly assign types independently to each line 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 (because 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

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

Mutations determine the types of some of the lines in the ASG even before we assign types to the lines at time $\beta = T$. Hence, we can prune 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, we can infer that at least one of the sampled individuals is fit. If all the lines end up in a deleterious mutation, 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 $\beta = T$ in the pruned ASG are assigned the unfit type. We use this idea in Section 4.2 to construct, for a given sample of the population at time $t = T$, branching–coalescing systems $\bar{\mathcal{G}} := (\bar{\mathcal{G}}(\beta))_{\beta \geq 0}$ and $\bar{\mathcal{G}}_T^\omega := (\bar{\mathcal{G}}_T^\omega(\beta))_{\beta \geq 0}$ in the annealed and quenched settings, respectively. Both processes have a cemetery state $\dagger$. The main feature of $\bar{\mathcal{G}}$ (resp. $\bar{\mathcal{G}}_T^\omega$) is that for any $\beta \geq 0$, the individuals in the sample are all unfit if and only if $\bar{\mathcal{G}} \neq \dagger$ (resp. $\bar{\mathcal{G}}_T^\omega \neq \dagger$) and all the lines present at time β in $\bar{\mathcal{G}}$ (resp. $\bar{\mathcal{G}}_T^\omega$) are unfit. We refer to $\bar{\mathcal{G}}$ and $\bar{\mathcal{G}}_T^\omega$ as the annealed and quenched killed ASGs (k-ASGs), respectively.

Moment dualities. We will now establish duality relations between the process X and the line-counting process of the k-ASG. For each $\beta \geq 0$, we denote by $R(\beta)$ the number of lines present in the annealed k-ASG at time β, with the convention that $R(\beta) = \dagger$ if $\bar{\mathcal{G}}(\beta) = \dagger$. The process $R := (R(\beta))_{\beta \geq 0}$, called the *annealed* line-counting process of the k-ASG, is a continuous-time Markov chain with values on $\mathbb{N}_0^\dagger := \mathbb{N}_0 \cup \{\dagger\}$ and infinitesimal generator matrix $Q_\dagger^\mu := \left(q_\dagger^\mu(i, j) \right)_{i,j \in \mathbb{N}_0^\dagger}$ defined via

$$q_\dagger^\mu(i, j) := \begin{cases} i(i-1) + i\theta\nu_1 & \text{if } j = i - 1, \\ \binom{i}{k}(\sigma_{i,k} + \sigma 1_{\{k=1\}}) & \text{if } j = i + k, \ i \geq k \geq 1, \\ i\theta\nu_0 & \text{if } j = \dagger, \\ -i(i - 1 + \theta + \sigma) - \int_{(0,1)} (1 - (1 - y)^i)\mu(\mathrm{d}y) & \text{if } j = i \in \mathbb{N}_0, \end{cases} \tag{2.6}$$

where the coefficients $\sigma_{m,k}$ are defined in Equation (2.5). All other entries are zero.

Similarly, for $T \in \mathbb{R}$ and a fixed environment $\omega \in \mathbb{D}^\star$, we denote by $R_T^\omega := (R_T^\omega(\beta))_{\beta \geq 0}$ the line-counting process associated to the quenched k-ASG $\bar{\mathcal{G}}_T^\omega$. The process R_T^ω, called the *quenched* line-counting process of the k-ASG, is a continuous-time (inhomogeneous) Markov process with values in $\mathbb{N}_0^\dagger$. It jumps from $i \in \mathbb{N}$ to $j \in \mathbb{N}_0^\dagger \setminus \{i\}$ at rate $q_\dagger^0(i, j)$, where $q_\dagger^0$ is the matrix defined in (2.6) with $\mu = 0$. In addition, at each time $\beta \geq 0$ with $\Delta\omega(T - \beta) > 0$, conditionally on $\{R_T^\omega(\beta -) = i\}$, $i \in \mathbb{N}$, we have $R_T^\omega(\beta) \sim i + \mathrm{Bin}(i, \Delta\omega(T - \beta))$. If $\theta > 0$ and $\nu_0 \in (0, 1)$, the states 0 and $\dagger$ are absorbing for R and R_T^ω.

Let J be a pure-jump subordinator with Lévy measure μ supported in $(0, 1)$. We write here X^J instead of X to stress the dependency of the (strong) solution to (1.3) on the environment

J. Similarly, we write X^ω for its quenched version (as introduced in Section 2.3). Since, in the annealed case, backward and forward environments have the same law, we can construct the line-counting process of the k-ASG as the strong solution to an SDE involving J and four other independent Poisson processes encoding the non-environmental events. We denote it by $(R^J(\beta))_{\beta \geq 0}$. The next result establishes a formal relation between X^J and R^J: a *reinforced moment duality*, which allows us to derive moment dualities in the annealed and quenched settings (see Figure 4 to visualize forward and backward processes in the same picture).

Theorem 2.3. (Reinforced, annealed, and quenched moment dualities.) *For all $x \in [0, 1]$, $n \in \mathbb{N}$, and $T \geq 0$, and any function $f \in \mathcal{C}^2([0, \infty))$ with compact support,*

$$\mathbb{E}[(1 - X^J(T))^n f(J(T)) \mid X^J(0) = x] = \mathbb{E}[(1 - x)^{R^J(T)} f(J(T)) \mid R^J(0) = n], \qquad (2.7)$$

with the convention $(1 - x)^\dagger = 0$ for all $x \in [0, 1]$. In particular, if $f = 1$ we recover the moment duality (we will often drop the superscript J when using this relation, unless we want to emphasize the dependency on J):

$$\mathbb{E}[(1 - X^J(T))^n \mid X^J(0) = x] = \mathbb{E}[(1 - x)^{R^J(T)} \mid R^J(0) = n]. \qquad (2.8)$$

For almost every (with respect to the law of J) environment $\omega \in \mathbb{D}^\star$,

$$\mathbb{E}\left[(1 - X^\omega(T))^n \mid X^\omega(0) = x\right] = \mathbb{E}\left[(1 - x)^{R_T^\omega(T-)} \mid R_T^\omega(0-) = n\right]. \qquad (2.9)$$

We prove (2.7) and (2.8) in Section 5.1. The proof of the quenched duality (2.9) is given in Section 6.1. Moreover, Theorem 7.1 extends (2.9) to any simple environment.

Remark 2.4. For $\theta = 0$, (2.8) is a particular case of [20, Lemma 2.14].

Asymptotic type composition. Assume now that $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$. In particular, the processes X and X^ω are not absorbed in $\{0, 1\}$. We will describe the asymptotic behavior of these processes using Theorem 2.3. The quenched case is particularly delicate, because for a given environment ω, $X^\omega(t)$ strongly depends on the environment in the recent past, and only weakly on the environment in the distant past (see Figure 3). Hence, unless ω is constant after some fixed time t_0 (i.e. ω has no jumps after t_0), $X^\omega(t)$ will not converge as $t \to \infty$ (see Remark 2.6 for the case of periodic environments). In contrast, for a given environment ω in $(-\infty, 0]$, we will see that $X^\omega(0)$, conditionally on $X^\omega(-\tau) = x$, converges in distribution as $\tau \to \infty$, and we will characterize its law; the setting is illustrated in Figure 5 (compare with Figure 3). To this end, for $n \in \mathbb{N}_0$ define

$$\pi_n := \mathbb{P}(\exists \beta \geq 0 : R(\beta) = 0 \mid R(0) = n),$$
$$\Pi_n(\omega) := \mathbb{P}(\exists \beta \geq 0 : R_0^\omega(\beta) = 0 \mid R_0^\omega(0-) = n), \qquad (2.10)$$

and set $\pi_\dagger := 0$ and $\Pi_\dagger(\omega) := 0$. Clearly, $\pi_0 = 1$ and $\Pi_0(\omega) = 1$.

Theorem 2.4. (Asymptotic type frequency.) *Assume that $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$.*

1. *The process X has a unique stationary distribution $\eta_X \in \mathcal{M}_1([0, 1])$, and letting $X(\infty)$ be a random variable distributed according to η_X, we have $X(t) \xrightarrow{(d)} X(\infty)$ as $t \to \infty$. Moreover, for all $n \in \mathbb{N}$,*

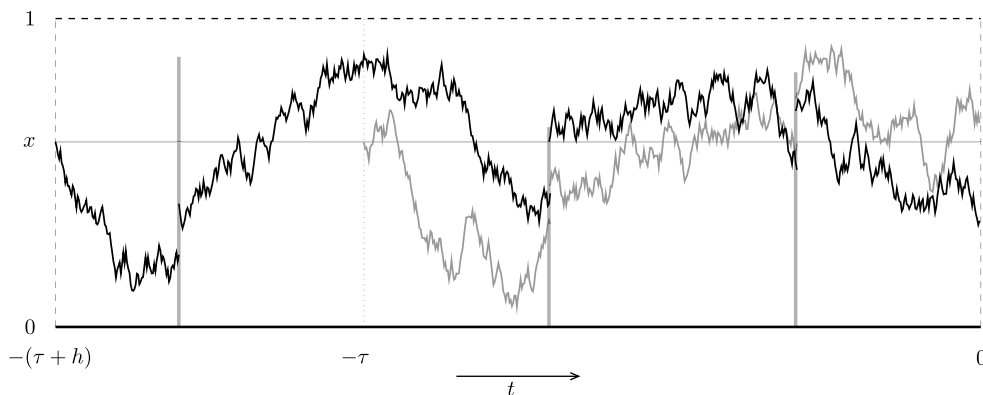$$\mathbb{E}\left[(1 - X(\infty))^n\right] = \pi_n, \qquad (2.11)$$

FIGURE 5. An illustration of two paths of X^ω: the black (resp. grey) path is defined in the interval $[-(\tau + h), 0]$ (resp. $[-\tau, 0]$) starting at $X^\omega(-(\tau + h)) = x$ (resp. $X^\omega(-\tau) = x$). The peaks of the environment ω are depicted as grey vertical lines.

and the absorption probabilities $(\pi_n)_{n\geq 0}$ satisfy

$$(\sigma + \theta + n - 1)\pi_n = \sigma\pi_{n+1} + (\theta\nu_1 + n - 1)\pi_{n-1}$$

$$+ \frac{1}{n}\sum_{k=1}^{n}\binom{n}{k}\sigma_{n,k}(\pi_{n+k} - \pi_n), \quad n \in \mathbb{N}, \qquad (2.12)$$

where the coefficients $\sigma_{n,k}$, $k \in [n]$, $n \in \mathbb{N}$, are defined in Equation (2.5).

2. *For almost every (with respect to the law of J) environment ω and for any $x \in (0, 1)$, the distribution of $X^\omega(0)$ conditionally on $\{X^\omega(-\tau) = x\}$ has a limit distribution $\mathcal{L}^\omega$ as $\tau \to \infty$, which does not depend on x. Moreover,*

$$\int_0^1 (1 - y)^n \mathcal{L}^\omega(\mathrm{d}y) = \Pi_n(\omega), \quad n \in \mathbb{N}, \qquad (2.13)$$

and the convergence of moments is exponential, i.e.

$$\left|\mathbb{E}\left[(1 - X^\omega(0))^n \mid X^\omega(-\tau) = x\right] - \Pi_n(\omega)\right| \leq e^{-\theta\nu_0\tau}, \quad n \in \mathbb{N}. \qquad (2.14)$$

The setting of Theorem 2.4(1) is illustrated in Figure 3, and its proof is given in Section 5.1; the setting of Theorem 2.4(2) is illustrated in Figure 5, and its proof is given in Section 6.1. Moreover, Theorem 7.2 extends Theorem 2.4(2) to any simple environment. A refinement of Theorem 2.4(2) is given in Theorem 7.3 for simple environments under additional conditions.

Remark 2.5. *Simpson's index* is a popular tool for describing population diversity. It represents the probability that two individuals chosen uniformly at random from the population have the same type. In our case it is given by $\mathrm{Sim}(t) := X(t)^2 + (1 - X(t))^2$. If the types represent different species, it gives a measure of bio-diversity. If the types represent two alleles of a gene for a given species, it measures *homozygosity*. As a consequence of Theorem 2.4, one can express the moments of $\mathrm{Sim}(\infty)$ in terms of the coefficients $(\pi_n)_{n\geq 0}$. In particular, we have

$$\mathbb{E}[\mathrm{Sim}(\infty)] = \mathbb{E}\left[X(\infty)^2 + (1 - X(\infty))^2\right] = 1 - 2\pi_1 + 2\pi_2.$$
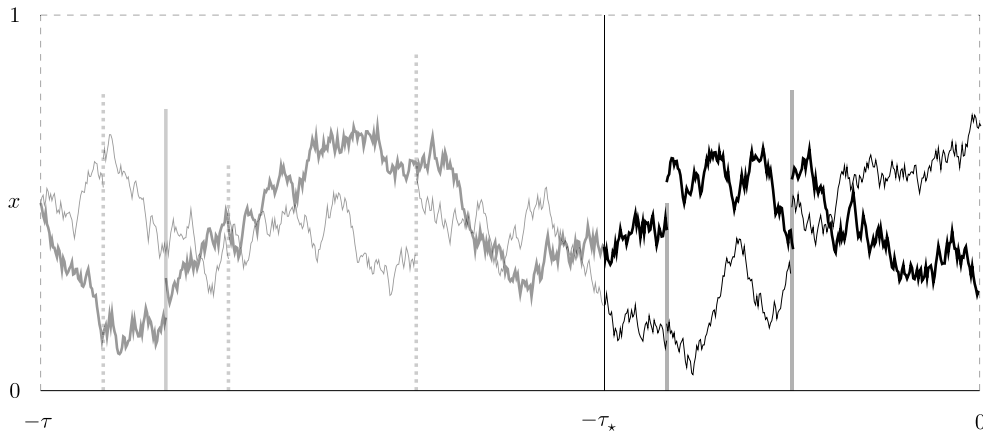
FIGURE 6. An illustration of two paths (thin and thick) of $X^{J \otimes_{\tau_\star} \omega}$ in $[-\tau, 0]$ starting at x. Both paths are subject to the same deterministic environment ω in $[-\tau_\star, 0]$; the vertical lines in $(-\tau_\star, 0]$ represent the peaks of ω. The environment in $[-\tau, -\tau_\star]$ is random and driven by J; the peaks of the realization of J giving rise to the thick (resp. thin) path in $[-\tau, -\tau_\star)$ are depicted as solid (resp. dotted) vertical lines.

Remark 2.6. If ω is a periodic environment on $[0, \infty)$ with period $T_p > 0$, the proof of Theorem 2.4(2) yields that, for any $x \in (0, 1)$ and $r \in [0, T_p)$, the distribution of $X^\omega(n T_p + r)$, conditionally on $\{X^\omega(0) = x\}$, has a limit distribution $\mathcal{L}_r^\omega$, when n goes to infinity, which is a function of ω and r and does not depend on x. Furthermore, $\mathcal{L}_r^\omega$ satisfies $\int_0^1 (1-y)^n \mathcal{L}_r^\omega(\mathrm{d}y) = \Pi_n(\omega_r)$, where ω_r is the periodic environment in $(-\infty, 0]$ defined by $\omega_r(t) := \omega(r + t + (\lfloor -t/T_p \rfloor + 1) T_p)$ for any $t \in (-\infty, 0]$. The convergence of moments is exponential as in (2.14).

Mixed environments. We present now an application illustrating the advantage of studying both quenched and annealed settings. We consider a population evolving from the distant past in a (stationary) random environment and analyze the effect of a recent perturbation of the environment on the type composition at present. To this end, we assume we only know the distribution of the environment before the perturbation.

In the absence of perturbations, the environment is given by a pure-jump subordinator J in $(-\infty, 0]$ with Lévy measure μ satisfying (2.4). The perturbation occurs in $(-\tau_\star, 0]$ (for some $\tau_\star > 0$) and is given by a deterministic environment ω. Let $X^{J \otimes_{\tau_\star} \omega}$ be the solution to (1.3) under the environment $J \otimes_{\tau_\star} \omega$, which coincides with J and ω in $(-\infty, -\tau_\star]$ and $(-\tau_\star, 0]$, respectively; see Figure 6 for an illustration. Recall that $(R_0^\omega(\beta))_{\beta \in [0, \tau_\star)}$ is the line-counting process associated to the quenched k-ASG (see Section 2.5). We are interested in the distribution of $X^{J \otimes_{\tau_\star} \omega}(0)$. The next result provides the moments of this random variable.

Proposition 2.1. *Assume that $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$. For any $\tau_\star > 0$, $n \in \mathbb{N}$, $x \in [0, 1]$, and almost every (with respect to the law of J) $\omega \in \mathbb{D}^\star$, we have*

$$\lim_{\tau \to \infty} \mathbb{E}\left[(1 - X^{J \otimes_{\tau_\star} \omega}(0))^n \mid X^{J \otimes_{\tau_\star} \omega}(-\tau) = x\right] = \mathbb{E}\left[\pi_{R_0^\omega(\tau_\star -)} \mid R_0^\omega(0-) = n\right]. \quad (2.15)$$

Proposition 2.1 is proved in Section 6.1; a refinement of this result is given for simple environments under the additional condition $\sigma = 0$ in Proposition 7.2.

2.6. Ancestral type via the pruned lookdown ASG

In this section we are interested in the type distribution at present of the individuals that will be successful in the long run. This distribution may differ substantially from the type composition at present and may show a bias towards the fit type.

Consider a sample of n individuals at some time T in the future and trace their ancestral lines using the ASG. We will see in Section 5.2 that the number of lines in the ASG is positive recurrent (see Lemma 5.2). Hence, the ASG has bottlenecks, and if T is sufficiently large, the n individuals share a common ancestor at time 0. Assigning types to the lines in the ASG at time 0 and propagating the types along using the pecking order, we determine the types in the sample as well as the true genealogy. In particular, we obtain the type of the common ancestor of the sample. What it means for T to be sufficiently large depends on n and on the realization of the ASG, but this dependency vanishes as $T \to \infty$. Because we are interested in the type of the individual that is successful in the long run, we can work under this limit consideration. In what follows we formalize this idea.

Consider a realization $\mathcal{G}_{[0,T]} := (\mathcal{G}(\beta))_{\beta \in [0,T]}$ of the annealed ASG in $[0, T]$ started with one line, representing an individual sampled at forward time T. If t denotes forward time, we set $\beta = T - t$ to denote the backward time (see Figure 4). For $\beta \in [0, T]$, let V_β be the set of lines present at time β in $\mathcal{G}_{[0,T]}$. Consider a function $c : V_T \to \{0, 1\}$ representing an assignment of types to the lines in V_T. Given $\mathcal{G}_{[0,T]}$ and c, we propagate types (forward in time) along the lines of $\mathcal{G}_{[0,T]}$, keeping track, at any time $\beta \in [0, T]$, of the true ancestor in V_T of each line in V_β. We denote by $a_c(\mathcal{G}_{[0,T]})$ the type of the ancestor in V_T of the single line in V_0. Assume that, under $\mathbb{P}_x$, c assigns independently to each line type 0 with probability x and type 1 with probability $1 - x$. The *annealed ancestral type distribution at time T* is

$$h_T(x) := \mathbb{P}_x(a_c(\mathcal{G}_{[0,T]}) = 0), \quad x \in [0, 1].$$

In the quenched setting we proceed in the same way, but using $\mathcal{G}^\omega_{[0,T]} := (\mathcal{G}^\omega_T(\beta))_{\beta \in [0,T]}$, the quenched ASG in $[0, T]$ in the environment ω of an individual sampled at time T, instead of $\mathcal{G}_{[0,T]}$. The *quenched ancestral type distribution at time T* is

$$h^\omega_T(x) := \mathbb{P}_x \left(a_c(\mathcal{G}^\omega_{[0,T]}) = 0 \right), \quad x \in [0, 1],$$

where, under $\mathbb{P}_x$, c assigns independently to each line present in $\mathcal{G}^\omega_{[0,T]}$ at time $\beta = T$ type 0 with probability x and type 1 with probability $1 - x$.

In the absence of mutations, the ancestor of the sampled individual is fit if and only if there is at least one fit line in the ASG having type 0 at time $\beta = T$. In the presence of mutations, determining the type of the ancestor is more involved. In [31] the ancestral type distribution was obtained for the null environment using the line-counting process of a pruned version of the ASG, called *the pruned lookdown ASG* (pLD-ASG). In Section 4.3 we generalize this construction to incorporate the effect of the environment. The main feature of the pLD-ASG is that the type of the ancestor at time $t = 0$ of the sampled individual at time $t = T$ is 0 if and only if there is at least one line in the pLD-ASG at time $\beta = T$ that has type 0 (see Lemma 4.3). Hence, $h_T(x)$ and $h^\omega_T(x)$ can be represented via the corresponding line-counting processes (which can easily be inferred from the description of the pLD-ASG given in Section 5.2).

The line-counting process of the annealed pLD-ASG, denoted by $L := (L(\beta))_{\beta \geq 0}$, is a continuous-time Markov chain with values on $\mathbb{N}$ and generator matrix $Q^\mu := (q^\mu(i, j))_{i,j \in \mathbb{N}}$ given by

$$q^\mu(i, j) := \begin{cases} i(i-1) + (i-1)\theta\nu_1 + \theta\nu_0 & \text{if } j = i-1, \\ i(\sigma + \sigma_{i,1}) & \text{if } j = i+1, \\ \binom{i}{k}\sigma_{i,k} & \text{if } j = i+k, \ i \geq k \geq 2, \\ \theta\nu_0 & \text{if } 1 \leq j < i-1, \\ -(i-1)(i+\theta) - i\sigma - \int_{(0,1)} (1 - (1-y)^i)\mu(\mathrm{d}y) & \text{if } j = i, \end{cases} \tag{2.16}$$

where $\sigma_{m,k}$ is defined in (2.5); all other entries are 0.

The pLD-ASG associated to $\omega \in \mathbb{D}^\star$ is well-defined and almost surely contains finitely many lines at any time; we show this in Section 5.2. The corresponding line-counting process $(L_T^\omega(\beta))_{\beta \geq 0}$ started at time T is a continuous-time (inhomogeneous) Markov process with values in $\mathbb{N}$. It jumps from $i \in \mathbb{N}$ to $j \in \mathbb{N} \setminus \{i\}$ at rate $q^0(i, j)$, where q^0 is the matrix defined in (2.16) with $\mu = 0$, and in addition, at each time $\beta \geq 0$ such that $\Delta\omega(T - \beta) > 0$, conditionally on $\{L_T^\omega(\beta-) = i\}$, $i \in \mathbb{N}$, we have $L_T^\omega(\beta) \sim i + \text{Bin}(i, \Delta\omega(T - \beta))$.

We now state the main result of this section, describing the asymptotic behavior of $h_T(x)$ and $h_T^\omega(x)$.

Theorem 2.5. (Ancestral type distribution.) *The following assertions hold:*

1. *The process L admits a unique stationary distribution η_L. Moreover, if $L(\infty)$ is a random variable distributed according to η_L, then $L(T) \xrightarrow{(d)} L(\infty)$ as $T \to \infty$. In particular, $h(x) := \lim_{T \to \infty} h_T(x)$ is well-defined, and*

$$h(x) = \sum_{n \geq 0} x(1-x)^n a_n, \tag{2.17}$$

where the coefficients $a_n := \mathbb{P}(L(\infty) > n)$, $n \in \mathbb{N}_0$, satisfy the following recursion (which is known as Fearnhead's recursion when $\mu = 0$):

$$(\sigma + \theta + n + 1) a_n = \sigma a_{n-1} + (\theta\nu_1 + n + 1) a_{n+1} + \frac{1}{n} \sum_{j=1}^n \gamma_{n+1,j} (a_{j-1} - a_j), \quad n \in \mathbb{N}, \tag{2.18}$$

where $\gamma_{i,j} := \sum_{k=i-j}^j \binom{j}{k}\sigma_{j,k}$ if $1 \leq j < i \leq 2j$ and $\gamma_{i,j} := 0$ otherwise.

2. *Assume that $\theta\nu_0 > 0$. For any $n \in \mathbb{N}$, the distribution of $L_T^\omega(T-)$ conditionally on $\{L_T^\omega(0-) = n\}$ has a limit distribution $\mu^\omega \in \mathcal{M}_1(\mathbb{N})$ as $T \to \infty$, which does not depend on n. In particular, $h^\omega(x) := \lim_{T \to \infty} h_T^\omega(x)$ is well-defined and*

$$h^\omega(x) = 1 - \sum_{n=1}^\infty \mu^\omega(\{n\})(1-x)^n. \tag{2.19}$$

Moreover, for any $x \in [0, 1]$ and $t > 0$,

$$\left| h^\omega(x) - h_T^\omega(x) \right| \leq 2e^{-\theta\nu_0 T}. \tag{2.20}$$

The proof of Theorem 2.5(1) is given in Section 5.2; Theorem 2.5(2) is proved in Section 6.2. Theorem 7.4 extends Theorem 2.5(2) to the case $\theta \nu_0 = 0$ for simple environments under additional conditions. A refinement of Theorem 2.5(2) is given in Theorem 7.5 for simple environments under additional conditions.

In the case $\theta = 0$, Theorem 2.5 yields the following result about the boundary behavior of X.

Corollary 2.1. (Accessibility of the boundaries.) *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, conditionally on $\{X(0) = x\}$, $X(T)$ converges almost surely as $T \to \infty$ to 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 2.7. Corollary 2.1 is not a direct consequence of [20, Theorem 3.2], whose statement does not cover SDEs with a diffusion term (the term $\sqrt{2X(t)(1 - X(t))}\mathrm{d}B(t)$). We close this section with an application of our results to the comparison of the (isolated) effects of the environment and of (genic) selection. To this end, we fix a non-zero measure μ on $(0, 1)$ satisfying (2.4) and we consider two models, both without mutations. The first model has selection parameter

$$\sigma = \sigma_\mu := \int_{(0,1)} y\mu(\mathrm{d}y), \tag{2.21}$$

and no environment (i.e. in (1.3) we take $S(t) := \sigma_\mu t$). The second one has selection parameter $\sigma = 0$ and an environment given by a subordinator with Lévy measure μ (i.e. in (1.3) we take $S(t) := J(t)$). We will use the superscript 'sel' (resp. 'env') to refer to the first (resp. second) model.

For $n \in \mathbb{N}$, set $\rho_n^{\mathrm{env}} := \mathbb{P}(L^{\mathrm{env}}(\infty) = n)$ and $\rho_n^{\mathrm{sel}} := \mathbb{P}(L^{\mathrm{sel}}(\infty) = n)$. Consider the probability generating functions

$$p^{\mathrm{env}}(z) := \sum_{n=1}^{\infty} \rho_n^{\mathrm{env}} z^n \quad \text{and} \quad p^{\mathrm{sel}}(z) := \sum_{n=1}^{\infty} \rho_n^{\mathrm{sel}} z^n, \quad z \in [0, 1].$$

Note that $p^{\mathrm{env}}(z) = 1 - h^{\mathrm{env}}(1 - z)$ and $p^{\mathrm{sel}}(z) = 1 - h^{\mathrm{sel}}(1 - z)$.

Proposition 2.2. *For any non-zero measure μ on $(0,1)$ satisfying (2.4) we have*

$$\rho_1^{\mathrm{env}} > \rho_1^{\mathrm{sel}} = \frac{\sigma_\mu}{e^{\sigma_\mu} - 1} \quad \text{and} \quad p^{\mathrm{env}}(z) \leq \frac{\rho_1^{\mathrm{env}}}{\rho_1^{\mathrm{sel}}} p^{\mathrm{sel}}(z) = \rho_1^{\mathrm{env}} \left(\frac{e^{\sigma_\mu z} - 1}{\sigma_\mu} \right), \quad z \in [0, 1].$$

In particular, there is $x_c \in (0, 1)$ such that, for $x \in [x_c, 1)$,

$$h^{\mathrm{env}}(x) = \mathbb{P}\left(\lim_{t \to \infty} X^{\mathrm{env}}(t) = 1 | X^{\mathrm{env}}(0) = x \right) < \mathbb{P}\left(\lim_{t \to \infty} X^{\mathrm{sel}}(t) = 1 | X^{\mathrm{sel}}(0) = x \right) = h^{\mathrm{sel}}(x).$$

Remark 2.8. As a consequence of Proposition 2.2 one recovers the classical result of Kimura [28],

$$h^{\mathrm{sel}}(x) = \frac{1 - e^{-\sigma_\mu x}}{1 - e^{-\sigma_\mu}}, \quad x \in [0, 1].$$

Remark 2.9. Consider a Wright–Fisher diffusion with no mutations and selection parameter σ, evolving in an environment with Lévy measure μ. The quantity σ_μ in (2.21) corresponds to the quantity $\alpha_\mathfrak{s}$ in [20]. As shown there, $\alpha_\mathfrak{s}$ is not sufficient to fully describe the strength of the environment; one also needs to know the *shape of rare selection*, which is defined as $\alpha^* := \int_{(0,1)} \log(1+y)\mu(\mathrm{d}y)/\alpha_\mathfrak{s}$. The joint action of weak selection and the environment is then described by the quantity $\alpha_{\mathrm{eff}} := \sigma + \alpha_\mathfrak{s}\alpha^*$, which is called the *effective strength of selection*. The main result in [20] establishes that both boundaries are accessible if and only if α_{eff} is smaller than a quantity β^* coding for neutral reproductions ($\beta^* = \infty$ in our case).

The proofs of Corollary 2.1 and Proposition 2.2 are given in Section 5.2.

3. Moran models and Wright–Fisher processes

This section is devoted to the proofs of Theorem 2.1 and Theorem 2.2 and other related results.

3.1. Results related to Section 2.1

Graphical representation. We start by making more precise the description of the graphical representation of the Moran model as an IPS. This will allow us to decouple the randomness of the model coming from the initial type configuration, the randomness coming from mutations and reproductions, and the randomness coming from the environment. Non-environmental events are as usual encoded via a family of independent Poisson processes $\Lambda := \{\lambda_i^0, \lambda_i^1, \{\lambda_{i,j}^\triangle, \lambda_{i,j}^\blacktriangle\}_{j \in [N] \setminus \{i\}}\}_{i \in [N]}$, where (a) for each $i,j \in [N]$ with $i \neq j$, $(\lambda_{i,j}^\triangle(t))_{t \in \mathbb{R}}$ and $(\lambda_{i,j}^\blacktriangle(t))_{t \in \mathbb{R}}$ are Poisson processes with rates σ_N/N and $1/N$, respectively, and (b) for each $i \in [N]$, $(\lambda_i^0(t))_{t \in \mathbb{R}}$ and $(\lambda_i^1(t))_{t \in \mathbb{R}}$ are Poisson processes with rates $\theta_N \nu_0$ and $\theta_N \nu_1$, respectively. We call Λ the *basic background*. The environment introduces a new independent source of randomness into the model, which we describe via the collection

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

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

Recall that in the graphical representation individuals are represented by horizontal lines at levels $i \in [N]$ (see Figure 1). The random appearance of selective and neutral arrows, circles, and crosses is prescribed by the background as follows. At the arrival times of $\lambda_{i,j}^\triangle$ (resp. $\lambda_{i,j}^\blacktriangle$), we draw selective (resp. neutral) arrows from level i to level j. At the arrival times of λ_i^0 (resp. λ_i^1), we draw an open circle (resp. a cross) at level i. Given an environment $\zeta := (t_k, p_k)_{k \in I}$ satisfying (2.1), we define, for each $k \in I$,

$$I_\zeta(k) := \{i \in [N] : U_i(t_k) \leq p_k\} \quad \text{and} \quad n_\zeta(k) := |I_\zeta(k)|,$$

and we draw, at time t_k, for each $i \in I_\zeta(k)$ a selective arrow from level i to level $\tau_{I_\zeta(k)}(t, i)$.

Continuity with respect to the environment. Now we embark on the proof of Theorem 2.1, which states the continuity of the type composition in a Moran model with respect to the

environment. The paths of the fit-counting process are considered as elements of $\mathbb{D}_{0,T}$, which is endowed with the J_1-Skorokhod topology, i.e. the topology induced by the Billingsley metric d_T^0 defined in (A.2). Recall also that the restriction of an environment to $[0, T]$ is described by means of a function in $\mathbb{D}_T^\star$ (see (2.3)), which is endowed with the topology induced by the metric $d_T^\star$ defined in (A.3).

Let us denote by $\mu_N(\omega)$ the law of $(Z_N^\omega(t))_{t \in [0,T]}$ (recall that $Z_N^\omega(t)$ is the number of fit individuals at time t in a Moran population of size N subject to environment ω). Theorem 2.1 states the continuity of the mapping $\omega \mapsto \mu_N(\omega)$, where the set of probability measures on $\mathbb{D}_{0,T}$ 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 ϱ_E defined in (A.6).

First, we get rid of the small jumps of the environment. To this end, we introduce the following notation. For $\delta > 0$ and $\omega \in \mathbb{D}_T^\star$, we define $\omega^\delta, \omega_\delta \in \mathbb{D}_T^\star$ via

$$\omega^\delta(t) := \sum_{u \in [0,t]: \Delta\omega(u) \geq \delta} \Delta\omega(u) \quad \text{and} \quad \omega_\delta(t) := \sum_{u \in [0,t]: \Delta\omega(u) < \delta} \Delta\omega(u). \tag{3.1}$$

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

$$d_t^\star(\omega, \omega^\delta) \leq \sum_{u \in [0,T]} |\Delta\omega(u) - \Delta\omega^\delta(u)| = \omega_\delta(T) \xrightarrow[\delta \to 0]{} 0.$$

In addition, for $\omega \in \mathbb{D}_T^\star$, $n \in \mathbb{N}$, and $\vec{r} := (r_i)_{i \in [n]} \in [0, T]^n$, we denote by $\mu_N^{\vec{r}}(\omega)$ the law of $(Z_N^\omega(r_i))_{i \in [n]}$, where $[0, N]^n$ is equipped with the distance d_1 defined via

$$d_1((x_i)_{i \in [n]}, (y_i)_{i \in [n]}) := \sum_{i \in [n]} |x_i - y_i|. \tag{3.2}$$

Proposition 3.1. *Let* $\omega \in \mathbb{D}_T^\star$. *Assume that for any* $\delta > 0$ *we have* $Z_N^{\omega^\delta}(0) = Z_N^\omega(0)$; *then*

$$\varrho_{[0,T]^n}(\mu_N^{\vec{r}}(\omega^\delta), \mu_N^{\vec{r}}(\omega)) \leq nN \omega_\delta(r_*) e^{\sigma_N r_* + \omega(r_*)}, \quad \forall \vec{r} \in [0, T]^n, n \in \mathbb{N}, \tag{3.3}$$

where $r_* := \max_{i \in [n]} r_i$. *Moreover,*

$$\varrho_{\mathbb{D}_{0,T}}(\mu_N(\omega^\delta), \mu_N(\omega)) \leq N\omega_\delta(T) e^{(1+\sigma_N)T + \omega(T)}. \tag{3.4}$$

In particular,

$$(Z_N^{\omega^\delta}(t))_{t \in [0,T]} \xrightarrow[\delta \to 0]{(d)} (Z_N^\omega(t))_{t \in [0,T]}.$$

Proof. For $\delta > 0$, we couple in $[0, T]$ a Moran model with parameters $(\sigma_N, \theta_N, \nu_0, \nu_1)$ and environment ω to a Moran model with parameters $(\sigma_N, \theta_N, \nu_0, \nu_1)$ and environment ω^δ (both of size N) by using the same initial type configuration, the same basic background, and the same environmental background. For any $t \in [0, T]$ and $a, b \in \{0, 1\}$, we denote by $Y_N^{a,b}(t)$ the number of individuals that at time t have type a under the environment ω and type b under the environment ω^δ. Clearly, we have

$$\left| Z_N^{\omega^\delta}(t) - Z_N^\omega(t) \right| = \left| Y_N^{1,0}(t) - Y_N^{0,1}(t) \right| \leq Y_N^{1,0}(t) + Y_N^{0,1}(t) := Y_N^{\neq}(t).$$

Note that $Y_N^{\neq}(t)$ is the number of individuals that have different types at time t under ω and ω^δ. In particular, we have $Y_N^{\neq}(t) \leq N$ almost surely. Let us assume that at time t a graphical element arises in the basic background, i.e. t is an arrival time of one of the Poisson processes in the family Λ. If the graphical element is a mutation, then $Y_N^{\neq}(t) \leq Y_N^{\neq}(t-)$. If the graphical element is a neutral arrow, we have

$$\mathbb{E}\left[Y_N^{\neq}(t) \mid Y_N^{\neq}(t-)\right] = Y_N^{\neq}(t-) + \frac{1}{N} Y_N^{\neq}(t-)(N - Y_N^{\neq}(t-)) - \frac{1}{N}(N - Y_N^{\neq}(t-))Y_N^{\neq}(t-)$$

$$= Y_N^{\neq}(t-).$$

If the graphical element is a selective arrow, then $Y_N^{\neq}(t)$ can increase by 1 only if the individual at the tail of the arrow has a different type at time t under ω and ω^δ. Thus

$$\mathbb{E}\left[Y_N^{\neq}(t) \mid Y_N^{\neq}(t-)\right] \leq \left(1 + \frac{1}{N}\right) Y_N^{\neq}(t-).$$

Now, let $0 \leq s < t \leq T$ and assume that there are neither jumps of ω^δ nor selective events in (s, t). In particular, in (s, t) only the population driven by ω is affected by the environment. Moreover, since neutral reproductions and mutations do not increase the expected value of $Y_N^{\neq}$, we obtain

$$\mathbb{E}\left[Y_N^{\neq}(t-) \mid Y_N^{\neq}(s)\right] \leq Y_N^{\neq}(s) + N \sum_{u \in (s,t)} \Delta\omega(u).$$

In addition, if at time t the environment ω^δ jumps (there are only finitely many of these jumps), then

$$\mathbb{E}\left[Y_N^{\neq}(t) \mid Y_N^{\neq}(t-)\right] \leq Y_N^{\neq}(t-)(1 + \Delta\omega(t)).$$

Let $0 \leq t_1 < \cdots < t_m \leq T$ be the jump times of ω^δ. The previous discussion yields

$$\mathbb{E}\left[Y_N^{\neq}(t_{i+1}) \mid Y_N^{\neq}(t_i)\right] \leq \mathbb{E}\left[\left(1 + \frac{1}{N}\right)^{K_i}\right] \left(Y_N^{\neq}(t_i) + N\epsilon_i(\delta)\right)(1 + \Delta\omega(t_{i+1})), \qquad (3.5)$$

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 $N\sigma_N(t_{i+1} - t_i)$. Hence,

$$\mathbb{E}\left[Y_N^{\neq}(t_{i+1}) \mid Y_N^{\neq}(t_i)\right] \leq e^{\sigma_N(t_{i+1} - t_i)}\left(Y_N^{\neq}(t_i) + N\epsilon_i(\delta)\right)(1 + \Delta\omega(t_{i+1})).$$

Iterating this formula and using that $Y_N^{\neq}(0) = 0$ yields

$$\mathbb{E}\left[Y_N^{\neq}(t)\right] \leq e^{\sigma_N t} N\omega_\delta(t) \prod_{t_i \leq t}(1 + \Delta\omega(t_i)) \leq N\,\omega_\delta(t)\, e^{\sigma_N t + \sum_{u \in [0,t]} \Delta\omega(u)}. \qquad (3.6)$$

Recall the definition of the space $\mathrm{BL}(E)$ in Appendix A.2. We equip $[0, N]^n$ with the distance d_1 defined in (3.2). For any $n \geq 1$ and $F \in \mathrm{BL}([0, N]^n)$, we have

$$\left|\int F\mathrm{d}\mu_N^{\vec{r}}(\omega^\delta) - \int F\mathrm{d}\mu_N^{\vec{r}}(\omega)\right| = \left|\mathbb{E}\left[F((Z_N^{\omega^\delta}(r_j))_{j \in [n]})\right] - \mathbb{E}\left[F((Z_N^{\omega}(r_j))_{j \in [n]})\right]\right|.$$

Hence, if $\|F\|_{\mathrm{BL}} \le 1$ (see (A.5) for the definition of $\|\cdot\|_{\mathrm{BL}}$) and we couple $Z_N^{\omega^\delta}(t)$ and $Z_N^\omega(t)$ as before, we get that

$$
\left| \mathbb{E}\left[F((Z_N^{\omega^\delta}(r_j))_{j\in[n]}) \right] - \mathbb{E}\left[F((Z_N^\omega(r_j))_{j\in[n]}) \right] \right|
$$

$$
\le \left| \mathbb{E}\left[d_1((Z_N^{\omega^\delta}(r_j))_{j\in[n]}, (Z_N^\omega(r_j))_{j\in[n]}) \right] \right| = \mathbb{E}\left[\sum_{j\in[n]} |Y_N^{\ne}(r_j)| \right]
$$

$$
\le \sum_{j\in[n]} N\omega_\delta(r_j) e^{\sigma_N r_j + \sum_{u\in[0,r_j]} \Delta\omega(u)},
$$

where the last bound comes from (3.6) applied at r_j, $j \in [n]$. Taking the supremum over all $F \in \mathrm{BL}([0, N]^n)$ with $\|F\|_{\mathrm{BL}} \le 1$ and using the definition of the distance $\varrho_{[0,N]^n}$ in (A.6), we get (3.3). Now, define $Y_N^*(t) := \sup_{u\in[0,t]} Y_N^{\ne}(u)$. If at time t a neutral event occurs, then

$$
\mathbb{E}[Y_N^*(t) \mid Y_N^*(t-)] \le \left(1 + \frac{1}{N} \right) Y_N^*(t-).
$$

Other events can be treated as before, leading to (3.5) with K_i being this time the number of selective and neutral events in (t_i, t_{i+1}). Hence, Equation (3.4) follows similarly to (3.3). The convergence of $Z_N^{\omega^\delta}$ towards Z_N^ω is a direct consequence of (3.4). $\qquad\square$

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

$$
(Z_N^{\omega_k}(t))_{t\in[0,T]} \xrightarrow[k\to\infty]{(d)} (Z_N^\omega(t))_{t\in[0,T]}. \tag{3.7}
$$

Proof. The proof consists of two parts. In the first part, we construct a time deformation $\lambda_k \in \mathcal{C}_T^\uparrow$ with suitable properties. In the second part, we compare $Z_N^{\omega_k} \circ \lambda_k$ and Z_N^ω under an appropriate coupling of the underlying Moran models.

Part 1: We assume, without loss of generality, that $d_T^\star(\omega_k, \omega) > 0$ for all $k \in \mathbb{N}$. Set $\epsilon_k := 2d_T^\star(\omega_k, \omega)$, so that $d_T^\star(\omega_k, \omega) < \epsilon_k$. By definition of the metric $d_T^\star$ in (A.3), there is $\varphi_k \in \mathcal{C}_T^\uparrow$ such that

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

where $\|\cdot\|_T^0$ is defined in (A.1). Denote by $r_1 < \cdots < r_n$ the consecutive jump times of ω in $[0, T]$. We assume without loss of generality that $0 < r_1 \le r_n < T$. The case where ω jumps at T can be reduced to the previous case, by extending ω_k, $k \in \mathbb{N}$, and ω to $[0, T + \varepsilon]$ as constants in $[T, T + \varepsilon]$. Set $\gamma_k := T\sqrt{e^{\epsilon_k} - 1}$. In the remainder of the proof we assume that k is sufficiently large, so that $\gamma_k \le \min_{i\in[n]_0}(r_{i+1} - r_i)/3$, where $r_0 := 0$ and $r_{n+1} := T$. This condition ensures that the intervals $I_i^k := [r_i - \gamma_k, r_i + \gamma_k]$, $i \in [n]$, are disjoint and contained in $[0, T]$. Now, define $\lambda_k : [0, T] \to [0, T]$ via the following:

(i) For $u \notin \cup_{i=1}^n I_i^k$: $\lambda_k(u) := u$.

(ii) For $u \in [r_i - \gamma_k, r_i]$: $\lambda_k(u) := \varphi_k(r_i) + m_i(u - r_i)$, where $m_i := (\varphi_k(r_i) - r_i + \gamma_k)/\gamma_k$.

(iii) 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 sufficiently large, so that $\epsilon_k < \log 2$, we can infer from $\|\varphi_k\|_T^0 \le \epsilon_k$ and from $\gamma_k = T\sqrt{e^{\epsilon_k} - 1}$ that m_i and $\bar{m}_i$ are positive. It is then straightforward to check that $\lambda_k \in \mathcal{C}_T^\uparrow$, $\lambda_k(I_i^k) = I_i^k$, $i \in [n]$, and that

$$\sum_{u \in [0,T]} |\Delta\omega(u) - \Delta\bar{\omega}_k(u)| \le \epsilon_k,$$

where $\bar{\omega}_k := \omega_k \circ \lambda_k$. Moreover, since $\|\varphi_k\|_T^0 \le \epsilon_k$, we infer that $\varphi_k(r_i) \in [e^{-\epsilon_k} r_i, e^{\epsilon_k} r_i]$. It follows that, for k sufficiently large,

$$1 - 2\sqrt{e^{\epsilon_k} - 1} \le m_i \le 1 + 2\sqrt{e^{\epsilon_k} - 1}, \quad i \in [n],$$

and the same holds for $\bar{m}_i$. Note that we can write $\lambda_k(t) = \int_0^t p_k(u)\mathrm{d}u$, with $p_k : [0, T] \mapsto \mathbb{R}$ taking only the values $(m_i)_{i \in [n]}$, $(\bar{m}_i)_{i \in [n]}$, and 1. In particular, we have $|p_k(u) - 1| \le 2\sqrt{e^{\epsilon_k} - 1}$ for all $u \in [0, T]$. Thus, for any $s, t \in [0, T]$ with $s \ne t$, the slope $(\lambda_k(s) - \lambda_k(t))/(s - t)$ belongs to $[1 - 2\sqrt{e^{\epsilon_k} - 1}, 1 + 2\sqrt{e^{\epsilon_k} - 1}]$. Therefore, for k sufficiently large, we have

$$\frac{\lambda_k(s) - \lambda_k(t)}{s - t}, \quad \frac{s - t}{\lambda_k(s) - \lambda_k(t)} \le 1 + 3\sqrt{e^{\epsilon_k} - 1}, \quad i \in [n].$$

Hence, using that $\log(1 + x) \le x$ for $x > -1$, we obtain for k sufficiently large

$$\|\lambda_k\|_T^0 \le 3\sqrt{e^{\epsilon_k} - 1}. \tag{3.8}$$

Part 2: For $\delta > 0$, we couple in $[0, T]$ a Moran model with parameters $(\sigma_N, \theta_N, \nu_0, \nu_1)$ and environment ω to a Moran model with parameters $(\sigma_N, \theta_N, \nu_0, \nu_1)$ and environment ω_k (both of size N) by using (1) the same initial type configuration and (2) the same basic background, and (3) by using in the second population the environmental background of the first one, time-changed by λ_k^{-1}. Under this coupling and by construction of the function λ_k, the Moran model associated to ω and the Moran model associated to ω_k (time-changed by λ_k) experience the same basic events out of the time intervals I_i^k. Moreover, at the times r_i, the success of simultaneous environmental reproductions is decided according to the same uniform random variables.

For $t \in [0, T]$, let $Y_N^{\ne}(t)$ be the number of individuals that have different types at time t for ω and at time $\lambda_k(t)$ for ω_k, and set $Y_N^*(t) := \sup_{u \in [0,t]} Y_N^{\ne}(u)$.

Consider the event $E_k :=$ there are no basic events in $\cup_{i \in [n]} I_i^k$, and note that

$$\mathbb{P}(E_k^c) \le n\left(1 - e^{-2N(1 + \sigma_N + \theta_N)\gamma_k}\right). \tag{3.9}$$

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

$$\mathbb{E}[Y_N^*(r_i + \gamma_k)1_{E_k}] \le \mathbb{E}[Y_N^*(r_i)1_{E_k}] + N \sum_{u \in (r_i, r_i + \gamma_k]} \Delta\bar{\omega}_k(u). \tag{3.10}$$

A similar argument yields

$$\mathbb{E}[Y_N^*(r_{i+1}-)1_{E_k}] \le \mathbb{E}[Y_N^*(r_{i+1} - \gamma_k)1_{E_k}] + N \sum_{u \in (r_{i+1} - \gamma_k, r_{i+1})} \Delta\bar{\omega}_k(u). \tag{3.11}$$

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

$$\mathbb{E}\left[Y_N^*(r_{i+1} - \gamma_k)1_{E_k}\right] \leq e^{(1+\sigma_N)(r_{i+1}-r_i)}\left(\mathbb{E}[Y_N^*(r_i+\gamma_k)1_{E_k}] + N\sum_{u\in J_i^k}\Delta\bar\omega_k(u)\right). \quad (3.12)$$

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

$$\mathbb{E}\left[Y_N^*(r_{i+1})\,1_{E_k}\right] \leq \mathbb{E}[Y_N^*(r_{i+1}-)\,1_{E_k}](1 + \Delta\omega(r_{i+1}) \wedge \Delta\bar\omega_k(r_{i+1}))$$
$$+ N|\Delta\omega(r_{i+1}) - \Delta\bar\omega_k(r_{i+1})|. \quad (3.13)$$

Using (3.10), (3.11), (3.12), and (3.13), we obtain

$$\mathbb{E}[Y_N^*(r_{i+1})\,1_{E_k}] \leq e^{(1+\sigma_N)(r_{i+1}-r_i)}\left[\mathbb{E}[Y_N^*(r_i)\,1_{E_k}]\right.$$
$$\left. + N\sum_{u\in(r_i,r_{i+1}]}|\Delta\omega(u) - \Delta\bar\omega_k(u)|\right](1 + \Delta\omega(r_{i+1})).$$

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

$$\mathbb{E}\left[Y_N^*(T)\,1_{E_k}\right] \leq N\epsilon_k e^{(1+\sigma_N)T + \sum_{u\in(0,T]}\Delta\omega(u)}. \quad (3.14)$$

Using (3.9), (3.14), (3.8), and the definition of d_T^0 in (A.2), we obtain for k large enough

$$\mathbb{E}\left[d_T^0(Z_N^\omega, Z_N^{\omega_k})\right] \leq 2nN\left(1 - e^{-2N(1+\sigma_N+\theta_N)\gamma_k}\right)$$
$$+ 3\sqrt{e^{\epsilon_k} - 1} \vee \left(N\epsilon_k\,e^{(1+\sigma_N)T + \sum_{u\in(0,T]}\Delta\omega(u)}\right).$$

The result follows from letting $k \to \infty$ and using that $\gamma_k \to 0$ and $\epsilon_k \to 0$ as $k \to \infty$. $\quad\square$

Proof of Theorem 2.1 *(continuity).* If ω has a finite number of jumps, the result follows directly from Proposition 3.2. In the general case, note that for any $\delta > 0$,

$$\varrho_{\mathbb{D}_{0,T}}(\mu_N(\omega_k), \mu_N(\omega)) \leq \varrho_{\mathbb{D}_{0,T}}\left(\mu_N(\omega_k), \mu_N\left(\omega_k^\delta\right)\right)$$
$$+ \varrho_{\mathbb{D}_{0,T}}\left(\mu_N\left(\omega_k^\delta\right), \mu_N\left(\omega^\delta\right)\right) + \varrho_{\mathbb{D}_{0,T}}(\mu_N(\omega^\delta), \mu_N(\omega)), \quad (3.15)$$

where ω^δ is as in (3.1) and, similarly, $\omega_k^\delta(t) := \sum_{u\in[0,t]:\Delta\omega_k(u)\geq\delta}\Delta\omega_k(u)$. Recall the definition of $d_T^\star$ in (A.3). We claim that, for any $\delta \in A_\omega := \{d > 0 : \Delta\omega(u) \neq d$ for any $u \in [0, T]\}$, we have

$$d_T^\star(\omega_k^\delta, \omega^\delta) \xrightarrow[k\to\infty]{} 0. \quad \text{(Claim 1)}$$

Assume that Claim 1 is true and let $\delta \in A_\omega$. Note that for any $\lambda \in \mathcal{C}_T^\uparrow$, we have

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

$$\leq |\omega_k(\lambda(T)) - \omega(T)| + |\omega(T) - \omega^\delta(T)| + |\omega^\delta(T) - \omega_k^\delta(\lambda(T))|$$

$$\leq d_T^0(\omega, \omega_k) + \omega_\delta(T) + d_T^0(\omega_k^\delta, \omega^\delta) \leq d_T^\star(\omega, \omega_k) + \omega_\delta(T) + d_T^\star(\omega_k^\delta, \omega^\delta),$$

where we used the definition of d_T^0 in (A.2) and then Lemma A.1. Combining this with Claim 1 and Proposition 3.1, we obtain

$$\limsup_{k \to \infty} \varrho_{\mathbb{D}_{0,T}}(\mu_N(\omega_k), \mu_N(\omega_k^\delta)) \leq N\omega_\delta(T)e^{(1+\sigma_N)T+\omega(T)}.$$

Proposition 3.2 and Claim 1 in turn yield $\limsup_{k \to \infty} \varrho_{\mathbb{D}_{0,T}}(\mu_N(\omega_k^\delta), \mu_N(\omega^\delta)) = 0$. Hence, letting $k \to \infty$ in (3.15) and using Proposition 3.1, we obtain

$$\limsup_{k \to \infty} \varrho_{\mathbb{D}_{0,T}}(\mu_N(\omega_k), \mu_N(\omega)) \leq 2N\omega_\delta(T)e^{(1+\sigma_N)T+\omega(T)}.$$

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 Claim 1. Let $\delta \in A_\omega$. Since $d_T^\star(\omega_k, \omega) \to 0$ as $k \to \infty$, we see from the definition of $d_T^\star$ in (A.3) that there exists $(\lambda_k)_{k\in\mathbb{N}}$ with $\lambda_k \in \mathcal{C}_T^\uparrow$ such that

$$\|\lambda_k\|_T^0 \xrightarrow[k\to\infty]{} 0 \quad \text{and} \quad \epsilon_k := \sum_{u \in [0,T]} |\Delta(\omega_k \circ \lambda_k)(u) - \Delta\omega(u)| \xrightarrow[k\to\infty]{} 0.$$

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,

$$\omega_k^\delta(\lambda_k(t)) - \omega^\delta(t) = \sum_{\substack{u \in [0,t] \\ \Delta\bar{\omega}_k(u) \geq \delta}} \Delta\bar{\omega}_k(u) - \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \geq \delta}} \Delta\omega(u) \leq \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \geq \delta-\epsilon_k}} \Delta\bar{\omega}_k(u) - \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \geq \delta}} \Delta\omega(u)$$

$$= \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \geq \delta-\epsilon_k}} (\Delta\bar{\omega}_k(u) - \Delta\omega(u)) + \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \in [\delta-\epsilon_k, \delta)}} \Delta\omega(u) \leq d_T^\star(\omega_k, \omega) + \sum_{\substack{u \in [0,T] \\ \Delta\omega(u) \in [\delta-\epsilon_k, \delta)}} \Delta\omega(u).$$

Similarly, we obtain

$$\omega^\delta(t) - \omega_k^\delta(\lambda_k(t)) = \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \geq \delta}} \Delta\omega(u) - \sum_{\substack{u \in [0,t] \\ \Delta\bar{\omega}_k(u) \geq \delta}} \Delta\bar{\omega}_k(u)$$

$$\leq \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) \in [\delta, \delta+\epsilon_k)}} \Delta\omega(u) + \sum_{\substack{u \in [0,t] \\ \Delta\bar{\omega}_k(u) \geq \delta}} (\Delta\omega(u) - \Delta\bar{\omega}_k(u)) \leq \sum_{\substack{u \in [0,T] \\ \Delta\omega(u) \in [\delta, \delta+\epsilon_k)}} \Delta\omega(u) + d_T^\star(\omega_k, \omega).$$

Thus, using the definition of d_T^0 in (A.2), we deduce that

$$d_T^0(\omega_k^\delta, \omega^\delta) \leq d_T^\star(\omega_k, \omega) + \sum_{\substack{u \in [0,T] \\ \Delta\omega(u) \in (\delta-\epsilon_k, \delta+\epsilon_k)}} \Delta\omega(u).$$

Since $\delta \in A_\omega$, letting $k \to \infty$ in the previous inequality yields $\lim_{k\to\infty} d_T^0(\omega_k^\delta, \omega^\delta) = 0$. Since ω^δ has a finite number of jumps, Claim 1 follows using Lemma A.1. $\qquad\square$

3.2. Results related to Section 2.3: the Wright–Fisher process as a large-population limit

We start this section by proving that the SDE (1.3) is well-posed.

Proposition 3.3. (Existence and uniqueness.) *Let σ, $\theta \geq 0$, ν_0, $\nu_1 \in [0, 1]$ with $\nu_0 + \nu_1 = 1$. Let J be a pure-jump subordinator with Lévy measure μ 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 (1.3) such that $X(0) = x_0$. Moreover, $X(t) \in [0, 1]$ for all $t \geq 0$.*

Remark 3.1. The Wright–Fisher diffusion defined via the SDE (1.3) with $\theta = 0$ corresponds to [20, Equation 10] with K_y, $y \in (0, 1)$, being a random variable that takes the value 1 with probability $1 - y$ and the value 2 with probability y.

Proof. We prove the existence and pathwise uniqueness of strong solutions to (1.3) via [32, Theorems 3.2 and 5.1]. To this end, we first extend (1.3) to an SDE on $\mathbb{R}$ and write it in the form of [32, Equation 2.1]. Note that by Lévy–Itô decomposition, the pure-jump subordinator J can be expressed as $J(t) = \int_{(0,1)} x \, N(t, \mathrm{d}x)$, where $N(\mathrm{d}s, \mathrm{d}x)$ is a Poisson random measure with intensity measure μ. We define the functions $a, b : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \times (0, 1) \to \mathbb{R}$ via

$$a(x) := \sqrt{2x(1-x)}, \quad b(x) := \sigma x(1-x) + \theta \nu_0 (1-x) - \theta \nu_1 x, \quad g(x, u) := x(1-x)u,$$

for $x \in [0, 1]$, $u \in (0, 1)$; $a(x) := 0$, $g(x, u) := 0$ for $x \notin [0, 1]$; $b(x) := \theta \nu_0$ for $x < 0$ and $b(x) := -\theta \nu_1$ for $x > 1$. Thus, Equation (1.3) can be extended to the following SDE on $\mathbb{R}$:

$$X(t) = X(0) + \int_0^t a(X(s)) \mathrm{d}B(s) + \int_0^t \int_{(0,1)} g(X(s-), u) N(\mathrm{d}s, \mathrm{d}u) + \int_0^t b(X(s)) \mathrm{d}s. \quad (3.16)$$

Thus, any solution $(X(t))_{t \geq 0}$ of (3.16) such that $X(t) \in [0, 1]$ for any $t \geq 0$ is a solution to (1.3), and vice versa. Note that the functions a, b, g are continuous. Moreover, $b = b_1 - b_2$, where

$$b_1(x) := \theta \nu_0 + \sigma(x \wedge 1)_+ \quad \text{and} \quad b_2(x) := \theta(x \wedge 1)_+ + \sigma(x \wedge 1)_+^2.$$

In addition, b_2 is non-decreasing. Thus, in order to apply [32, Theorems 3.2 and 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 proving that $x \mapsto b_1(x) + \int_{(0,1)} g(x, u) \mu(\mathrm{d}u)$ is Lipschitz continuous. In fact, a straightforward calculation shows that

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

and hence (3.a) follows. Condition (3.b) amounts to proving that $x \mapsto a(x)$ is 1/2-Hölder, which is already shown in the proof of [19, Lemma 3.6]. Therefore, [32, Theorem 3.2] yields the pathwise uniqueness for (3.16). Condition (5.a) follows from the fact that the functions a, b, $x \mapsto \int_{(0,1)} g(x, u)^2 \mu(\mathrm{d}u)$, and $x \mapsto \int_{(0,1)} g(x, u) \mu(\mathrm{d}u)$ are bounded on $\mathbb{R}$. Hence, [32, Theorem 5.1] ensures the existence of a strong solution to (3.16). It remains to show that any solution to (3.16) 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, Proposition 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, Proposition 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. $\quad \square$

Lemma 3.1. *The solution to the SDE* (1.3) *is a Feller process with generator $\mathcal{A}$ satisfying, for all $f \in \mathcal{C}^2([0, 1], \mathbb{R})$,*

$$\mathcal{A}f(x) = x(1-x)f''(x) + (\sigma x(1-x) + \theta \nu_0(1-x) - \theta \nu_1 x)f'(x)$$
$$+ \int_{(0,1)} (f(x + x(1-x)u) - f(x))\, \mu(\mathrm{d}u).$$

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

Proof. Since pathwise uniqueness implies weak uniqueness (see [6, Theorem 1]), we infer from [30, Corollary 2.16] that the martingale problem associated to $\mathcal{A}$ in $\mathcal{C}^2([0, 1])$ is well-posed. Moreover, an inspection of the proof shows that this is also true in $\mathcal{C}^\infty([0, 1])$. Using [37, Proposition 2.2], we deduce that X is Feller. The fact that $\mathcal{C}^\infty([0, 1])$ is a core then follows from [37, Theorem 2.5]. □

Now we will prove the first part of the main result of Section 2.3, i.e. the annealed convergence of a sequence of Moran models towards the solution to the SDE (1.3).

Proof of Theorem 2.2(1) *(annealed convergence.)* Let $\mathcal{A}_N^*$ and $\mathcal{A}$ be the infinitesimal generators of the processes $(X_N(t))_{t \geq 0}$ and $(X(t))_{t \geq 0}$, respectively. Note that $(X_N(t))_{t \geq 0}$ has state space

$$E_N := \{k/N : k \in [N]_0\}. \tag{3.17}$$

We will prove that, for all $f \in \mathcal{C}^\infty([0, 1], \mathbb{R})$,

$$\sup_{x \in E_N} |\mathcal{A}_N^* f(x) - \mathcal{A}f(x)| \xrightarrow[N \to \infty]{} 0. \tag{Claim 2}$$

Provided Claim 2 is true, since X is Feller and $\mathcal{C}^\infty([0, 1], \mathbb{R})$ is an operator core for $\mathcal{A}$ (see Lemma 3.1), the result follows from applying [24, Theorem 17.25]. Thus, it remains to prove Claim 2. To this end, we write $\mathcal{A}$ as $\mathcal{A}^1 + \mathcal{A}^2 + \mathcal{A}^3 + \mathcal{A}^4$, where

$$\mathcal{A}^1 f(x) := x(1-x)f''(x),$$

$$\mathcal{A}^2 f(x) := (\sigma x(1-x) + \theta \nu_0(1-x) - \theta \nu_1 x)f'(x),$$

$$\mathcal{A}^3 f(x) := \int_{(0,\varepsilon_N)} (f(x + x(1-x)u) - f(x))\, \mu(\mathrm{d}u),$$

$$\mathcal{A}^4 f(x) := \int_{(\varepsilon_N,1)} (f(x + x(1-x)u) - f(x))\, \mu(\mathrm{d}u).$$

We also write $\mathcal{A}_N^* = \mathcal{A}_N^1 + \mathcal{A}_N^2 + \mathcal{A}_N^3 + \mathcal{A}_N^4$, where

$$\mathcal{A}_N^1 f(x) := N^2 x(1-x) \left[\Delta_{\frac{1}{N}} f(x) + \Delta_{-\frac{1}{N}} f(x) \right],$$

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

$$\mathcal{A}_N^3 f(x) := \int_{(0,\varepsilon_N)} \mathbb{E}\left[f(x + \xi_N(x, u)) \right] - f(x)\mu(\mathrm{d}u),$$

$$\mathcal{A}_N^4 f(x) := \int_{(\varepsilon_N,1)} \mathbb{E}\left[f(x + \xi_N(x, u)) \right] - f(x)\mu(\mathrm{d}u),$$

where $\Delta_h f(x) := f(x + h) - f(x)$, and $\xi_N(x, u) := \mathcal{H}(N, N(1 - x), B_{Nx}(u))/N$, with $\mathcal{H}(N, N(1 - x), k) \sim \mathrm{Hyp}(N, N(1 - x), k)$, and $B_{Nx}(u) \sim \mathrm{Bin}(Nx, u)$ being independent; $\varepsilon_N > 0$ will be chosen later in an appropriate way.

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

$$\sup_{x \in E_N} |\mathcal{A}_N^* f(x) - \mathcal{A} f(x)| \le \sum_{i=1}^4 \sup_{x \in E_N} |\mathcal{A}_N^i f(x) - \mathcal{A}^i f(x)|. \tag{3.18}$$

Taylor expansions of order three around x for $f(x + 1/N)$ and $f(x - 1/N)$ yield

$$\sup_{x \in E_N} |\mathcal{A}_N^1 f(x) - \mathcal{A}^1 f(x)| \le \frac{\|f'''\|_\infty}{3N}. \tag{3.19}$$

Similarly, the triangle inequality and appropriate Taylor expansions of order two yield

$$\sup_{x \in E_N} |\mathcal{A}_N^2 f(x) - \mathcal{A}^2 f(x)| \le \frac{(N\sigma_N + N\theta_N)\|f''\|_\infty}{2N} + (|\sigma - N\sigma_N| + |\theta - N\theta_N|)\|f'\|_\infty. \tag{3.20}$$

Since $N\sigma_N \to \sigma$ and $N\theta_N \to \theta$, the right-hand side in (3.20) converges to 0 as $N \to \infty$. In addition, since $\mathbb{E}[\xi_N(x, u)] = x(1 - x)u$, we have

$$|\mathcal{A}_N^3 f(x)| \le \|f'\|_\infty \int_{(0, \varepsilon_N)} u\mu(\mathrm{d}u), \quad x \in [0, 1],$$

and hence,

$$\sup_{x \in E_N} |\mathcal{A}_N^3 f(x) - \mathcal{A}^3 f(x)| \le 2\|f'\|_\infty \int_{(0, \varepsilon_N)} u\mu(\mathrm{d}u). \tag{3.21}$$

For the last term, note first that

$$\left| \mathbb{E}\left[f(x + \xi_N(x, u)) - f(x + x(1 - x)u) \right] \right| \le \|f'\|_\infty \mathbb{E}\left[|\xi_N(x, u) - x(1 - x)u| \right]$$
$$\le \|f'\|_\infty \sqrt{\mathbb{E}\left[(\xi_N(x, u) - x(1 - x)u)^2 \right]} \le \|f'\|_\infty \sqrt{\frac{u}{N}},$$

where in the last inequality we used that

$$\mathbb{E}\left[(\xi_N(x, u) - x(1 - x)u)^2 \right] = \frac{x(1 - x)^2 u(1 - u)}{N} + \frac{Nx^2(1 - x)u^2}{N^2(N - 1)} \le \frac{u}{N}, \tag{3.22}$$

which is obtained from standard properties of the hypergeometric and binomial distributions. Hence,

$$\sup_{x \in E_N} |\mathcal{A}_N^4 f(x) - \mathcal{A}^4 f(x)| \le \|f'\|_\infty \int_{(\varepsilon_N, 1)} \sqrt{\frac{u}{N}} \mu(\mathrm{d}u) \le \frac{\|f'\|_\infty}{\sqrt{N}\varepsilon_N} \int_{(\varepsilon_N, 1)} u \mu(\mathrm{d}u). \tag{3.23}$$

Now, choose $\varepsilon_N := 1/\sqrt{N}$. Since $\int_{(0, 1)} u\mu(\mathrm{d}u) < \infty$, Claim 2 follows from plugging (3.19), (3.20), (3.21), and (3.23) into (3.18) and letting $N \to \infty$. $\qquad\square$

Before embarking on the proof of the second part of Theorem 2.2, we prove the following estimates for the Moran model with null environment.

Lemma 3.2. *For any $x \in E_N$ (see (3.17) for its definition) and $t \ge 0$, we have*

$$\mathbb{E}\left[\left(X_N^{\mathbf{0}}(t) - x \right)^2 \mid X_N^{\mathbf{0}}(0) = x \right] \le \left(\frac{1}{2} + N(\sigma_N + 3\theta_N) \right) t,$$

and

$$-N\theta_N\nu_1 t \leq \mathbb{E}\left[X_N^{\mathbf{0}}(t) - x \mid X_N^{\mathbf{0}}(0) = x\right] \leq N(\sigma_N + \theta_N\nu_0)t.$$

Proof. Fix $x \in E_N$ and consider the functions $f_x, g_x : E_N \to [0, 1]$ defined via $f_x(z) := (z - x)^2$ and $g_x(z) := z - x$, $z \in E_N$. The process $X_N^{\mathbf{0}}$ is a Markov chain with generator $\mathcal{A}_N^{\star,0} := \mathcal{A}_N^1 + \mathcal{A}_N^2$, where $\mathcal{A}_N^1$ and $\mathcal{A}_N^2$ are defined in the proof of Theorem 2.2. Moreover, for every $z \in E_N$, we have

$$\mathcal{A}_N^{\star,0} f_x(z) = 2z(1 - z)$$
$$+ N\left[(\sigma_N z + \theta_N\nu_0)(1 - z)\left(2(z - x) + \frac{1}{N}\right) + \theta_N\nu_1 z\left(2(x - z) + \frac{1}{N}\right)\right]$$
$$\leq \frac{1}{2} + N\left[3\left(\frac{\sigma_N}{4} + \theta_N\nu_0\right) + 3\theta_N\nu_1\right] \leq \frac{1}{2} + N(\sigma_N + 3\theta_N),$$

and

$$\mathcal{A}_N^{\star,0} g_x(z) = N\left[(\sigma_N z + \theta_N\nu_0)(1 - z) - \theta_N\nu_1 z\right] \in [-N\theta_N\nu_1, N(\sigma_N + \theta_N\nu_0)].$$

Hence, Dynkin's formula applied to X_N^0 with f_x leads to

$$\mathbb{E}\left[\left(X_N^{\mathbf{0}}(t) - x\right)^2 \mid X_N^{\mathbf{0}}(0) = x\right] = \int_0^t \mathbb{E}\left[\mathcal{A}_N^{\star,0} f_x(X_N^{\mathbf{0}}(s)) \mid X_N^{\mathbf{0}}(0) = x\right] ds$$
$$\leq \left(\frac{1}{2} + N(\sigma_N + 3\theta_N)\right) t.$$

Similarly, applying Dynkin's formula to X_N^0 with g_x, we obtain

$$\mathbb{E}\left[X_N^{\mathbf{0}}(t) - x \mid X_N^{\mathbf{0}}(0) = x\right] = \int_0^t \mathbb{E}\left[\mathcal{A}_N^{\star,0} g_x(X_N^{\mathbf{0}}(s)) \mid X_N^{\mathbf{0}}(0) = x\right] ds$$
$$\in [-N\theta_N\nu_1 t, N(\sigma_N + \theta_N\nu_0)t],$$

which ends the proof. $\qquad\square$

Proposition 3.4. (Quenched tightness.) *Assume that $N\sigma_N \to \sigma$ and $N\theta_N \to \theta$ as $N \to \infty$. For any $\omega \in \mathbb{D}^\star$, the sequence $(X_N^\omega)_{N \geq 1}$ is tight.*

Proof. Let $(\mathcal{F}_s^N)_{s \geq 0}$ denote the natural filtration associated to the process X_N^ω. To prove the tightness of the sequence $(X_N^\omega)_{N \geq 1}$, we use [5, Theorem 1]. For this we need to show that the following conditions hold:

(A1) For each $T, \varepsilon > 0$, there exists a compact set $K \subset \mathbb{R}$ such that

$$\liminf_{N \to \infty} \mathbb{P}\left(X_N^\omega(t) \in K, \ \forall t \leq T\right) \geq 1 - \varepsilon.$$

(A2) There exist $\alpha > 0$ and non-decreasing, càdlàg processes F_N, F such that F_N is $\mathcal{F}_0$-measurable, $F_N \overset{(d)}{\Longrightarrow} F$, and for any $N \geq 1$ and every $0 \leq s \leq t$,

$$\mathbb{E}\left[1 \wedge |X_N^\omega(t) - X_N^\omega(s)|^\alpha\right] \leq F_N(t) - F_N(s).$$

Since, for all $t \geq 0$ and $N \geq 1$, $X_N^\omega(t) \in E_N \subset [0, 1]$ (see (3.17) for the definition of E_N), Condition (A1) is satisfied. Now, we claim that there are constants $c, C > 0$, independent of N, such that

$$\mathbb{E}\left[(X_N^\omega(t) - X_N^\omega(s))^2 \mid \mathcal{F}_s^N\right] \leq c \sum_{u \in [s,t]} \Delta\omega(u) + C(t - s), \quad \text{for all } 0 \leq s \leq t. \quad \text{(Claim 3)}$$

If Claim 3 is true, then Condition (A2) is satisfied with $\alpha = 2$ and $F_N(t) = F(t) = c \sum_{u \in [0,t]} \Delta\omega(u) + Ct$, and the result follows from [5, Theorem 1]. The rest of the proof is devoted to proving Claim 3.

For $x \in E_N$ and $t \geq 0$, we set $\psi_x(\omega, t) := \mathbb{E}_x[(X_N^\omega(t) - x)^2]$. For $s \geq 0$, we set $\omega_s(\cdot) := \omega(s + \cdot)$. From the definition of X_N^ω, it follows that for any $0 \leq s < t$,

$$\mathbb{E}\left[(X_N^\omega(t) - X_N^\omega(s))^2 \Big| \mathcal{F}_s^N\right] = \psi_{X_N^\omega(s)}(\omega_s, t - s). \quad (3.24)$$

Let $0 \leq s < t$. We split the proof of Claim 3 into three cases.

Case 1: ω has no jumps in $(s, t]$. In particular, ω_s has no jumps in $(0, t - s]$. Hence, restricted to $[0, t - s]$, $X_N^{\omega_s}$ has the same distribution as $X_N^{\mathbf{0}}$. Using Lemma 3.2 with $x = X_N^\omega(s)$ and plugging the result into (3.24), we infer that Claim 3 holds for any $c \geq 1$ and $C \geq C_1 := 1/2 + \sup_{N \in \mathbb{N}} (N(\sigma_N + 3\theta_N))$.

Case 2: ω has n jumps in $(s, t]$. Let $t_1, \ldots, t_n \in (s, t]$ be the jump times of ω in $(s, t]$ in increasing order. We set $t_0 := s$ and $t_{n+1} = t$. For any $i \in [n + 1]$ and any $r \in (t_{i-1}, t_i)$, ω has no jumps in $(t_{i-1}, r]$. In particular, $(t_{i-1}, r]$ falls into Case 1. Using Claim 3 in $(t_{i-1}, r]$ and letting $r \to t_i$, we obtain

$$\mathbb{E}\left[(X_N^\omega(t_i-) - X_N^\omega(t_{i-1}))^2 \mid \mathcal{F}_{t_{i-1}}^N\right] \leq C_1(t_i - t_{i-1}).$$

Moreover,

$$\mathbb{E}\left[(X_N^\omega(t_i) - X_N^\omega(t_i-))^2 \mid \mathcal{F}_{t_i-}^N\right] \leq \mathbb{E}\left[\left(\frac{B_N(\Delta\omega(t_i))}{N}\right)^2\right] \leq \Delta\omega(t_i),$$

where $B_N(\Delta\omega(t_i)) \sim \text{Bin}(N, \Delta\omega(t_i))$. Using the two previous inequalities and the tower property of the conditional expectation, we get

$$\mathbb{E}\left[(X_N^\omega(t_i) - X_N^\omega(t_{i-1}))^2 \mid \mathcal{F}_s^N\right] \leq 2C_1(t_i - t_{i-1}) + 2\Delta\omega(t_i). \quad (3.25)$$

Now, note that

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

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

Using Equation (3.25), we see that

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

Moreover, we have

$$\mathbb{E}\left[(X_N^\omega(t_{i+1}) - X_N^\omega(t_i))(X_N^\omega(t_i) - X_N^\omega(s)) \mid \mathcal{F}_{t_i}^N\right] = \varphi_{X_N^\omega(s), X_N^\omega(t_i)}(\omega_{t_i}, t_{i+1} - t_i),$$

where for $x, y \in E_N$ and $t \geq 0$ we set $\varphi_{x,y}(\omega, t) := (y - x)\mathbb{E}[X_N^\omega(t) - y \mid X_N^\omega(0) = y]$. Since, for any $r \in (t_i, t_{i+1})$, ω_{t_i} has no jumps in $(0, r - t_i]$, we can use Lemma 3.2 to infer that for any $x, y \in E_N$,

$$\varphi_{x,y}(\omega_{t_i}, r - t_i) \leq N((\sigma_N + \theta_N v_0) \vee \theta_N v_1)(r - t_i).$$

Note that $(y - x)\mathbb{E}_y[X_N^{\omega_{t_i}}(t_{i+1} - t_i) - X_N^{\omega_{t_i}}((t_{i+1} - t_i)-)] \leq \Delta\omega(t_{i+1})$. Hence, letting $r \to t_{i+1}$, we get

$$\varphi_{x,y}(\omega_{t_i}, t_{i+1} - t_i) \leq N((\sigma_N + \theta_N v_0) \vee \theta_N v_1)(t_{i+1} - t_i) + \Delta\omega(t_{i+1}),$$

for any $x, y \in E_N$, hence in particular for $x = X_N^\omega(s)$ and $y = X_N^\omega(t_i)$. Altogether, we obtain

$$\mathbb{E}\left[(X_N^\omega(t) - X_N^\omega(s))^2 \mid \mathcal{F}_s^N\right] \leq C_2(t - s) + 3\sum_{i=1}^n \Delta\omega(t_i),$$

where $C_2 := 2C_1 + \sup_{N \in \mathbb{N}} N((\sigma_N + \theta_N v_0) \vee \theta_N v_1)$. Hence, Claim 3 holds for any $C \geq C_2$ and $c \geq 3$.

Case 3: ω has infinitely many jumps in $(s, t]$. For any δ, we consider ω^δ as in (3.1) and couple the processes X_N^ω and $X_N^{\omega^\delta}$ as in the proof of Proposition 3.1. Note that ω^δ has only a finite number of jumps in any compact interval; thus ω^δ falls into Case 2. Moreover, we have

$$\psi_x(\omega, t) \leq 2\mathbb{E}_x[(X_N^\omega(t) - X_N^{\omega^\delta}(t))^2] + 2\mathbb{E}_x\left[(X_N^{\omega^\delta}(t) - x)^2\right]$$

$$\leq 2\mathbb{E}_x[|X_N^\omega(t) - X_N^{\omega^\delta}(t)|] + 2\mathbb{E}_x\left[(X_N^{\omega^\delta}(t) - x)^2\right]$$

$$\leq 2e^{N\sigma_N t + \omega(t)} \sum_{\substack{u \in [0,t] \\ \Delta\omega(u) < \delta}} \Delta\omega(u) + 2\mathbb{E}_x\left[(X_N^{\omega^\delta}(t) - x)^2\right],$$

where in the last inequality we used Proposition 3.1. Now, using Claim 3 for $X_N^{\omega^\delta}$ and the previous inequality, we obtain

$$\mathbb{E}\left[(X_N^\omega(t) - X_N^\omega(s))^2 \mid \mathcal{F}_s^N\right] \leq e^{N\sigma_N(t-s) + \omega(t-s)} \sum_{\substack{u \in [s,t] \\ \Delta\omega(u) < \delta}} \Delta\omega(u) + 2C_2(t - s) + 6\sum_{u \in [s,t]} \Delta\omega(u).$$

We let $\delta \to 0$ and conclude that Claim 3 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 2.2(2) (quenched convergence). Let $B := (B(t))_{t \geq 0}$ be a standard Brownian motion. We denote by X^0 the unique strong solution to (1.3) associated to B and the null environment. Theorem 2.2(1) implies in particular that X_N^0 converges to X^0 as $N \to \infty$.

Now, assume that $\omega \neq \mathbf{0}$ is simple. We denote by T_ω the (discrete but possibly infinite) set of jump times of ω in $(0, \infty)$. Moreover, for $0 < i < |T_\omega| + 1$, we denote by $t_i := t_i(\omega) \in T_\omega$ the time of the ith jump of ω. We set $t_0 := 0$ and $t_{|T_\omega|+1} := \infty$. We need to prove that

$$(X_N^\omega(t))_{t \geq 0} \xrightarrow[N \to \infty]{(d)} (X^\omega(t))_{t \geq 0}.$$

Recall that the process X^ω starting at x_0 can be constructed as follows:

(i) (i)$X^\omega(0) = x_0$.

(ii) For $i \in \mathbb{N}$ with $i \leq |T_\omega| + 1$, the restriction of X^ω to the interval (t_{i-1}, t_i) is given by a version of $X^{\mathbf{0}}$ started at $X^\omega(t_{i-1})$.

(iii) For $0 < i < |T_\omega| + 1$, conditionally on $X^\omega(t_i -)$,

$$X^\omega(t_i) = X^\omega(t_i -) + X^\omega(t_i -)(1 - X^\omega(t_i -))\Delta\omega(t_i).$$

Since the sequence $(X_N^\omega)_{N \in \mathbb{N}}$ is tight (see Proposition 3.4), 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

$$((X_N^\omega(t))_{t \in I}, X_N^\omega(t_i -)) \xrightarrow[N \to \infty]{(d)} ((X^\omega(t))_{t \in I}, X^\omega(t_i -)).$$

For $i = |T_\omega| + 1 < \infty$ we remove the components $X_N^\omega(t_i -)$ and $X^\omega(t_i -)$ since they do not make sense. Since $X_N^\omega(t_1 -) = X_N^{\mathbf{0}}(t_1)$ and $X^\omega(t_1 -) = X^{\mathbf{0}}(t_1)$ almost surely, the result for $i = 1$ follows from Theorem 2.2(1). 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 by an analogous argument.

Let $F : [0, 1]^{k+1} \to \mathbb{R}$ be a Lipschitz function with $\|F\|_{\mathrm{BL}} \leq 1$ (see (A.5) for the definition of $\|\cdot\|_{\mathrm{BL}}$). Note that

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

where for $x \in E_N$ (see (3.17) for the definition of E_N) and $u \in (0, 1)$, we let $\xi_N(x, u) := \mathcal{H}(N, N(1 - x), B_{Nx}(u))/N$ with $\mathcal{H}(N, N(1 - x), k) \sim \mathrm{Hyp}(N, N(1 - x), k)$, $k \in [N]_0$, and $B_{Nx}(u) \sim \mathrm{Bin}(Nx, u)$ being independent of each other and independent of X_N^ω. Now, set

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

Using that $\|F\|_{\mathrm{BL}} \leq 1$ and (3.22), we see that $|D_N| \leq \sqrt{\Delta\omega(t_i)/N} \to 0$ as $N \to \infty$. Therefore, the induction hypothesis yields

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

Therefore,

$$\left(\left(X_N^\omega \left(s_j \right) \right)_{j=1}^k, X_N^\omega(t_i) \right) \xrightarrow[N\to\infty]{(d)} \left(\left(X^\omega \left(s_j \right) \right)_{j=1}^k, X^\omega(t_i) \right). \tag{3.26}$$

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

$$H_N(z, x) := \mathbb{E}_x[G(z, x, (X_N^{\mathbf{0}}(r_j - t_i))_{j=1}^m, X_N^{\mathbf{0}}(t_{i+1} - t_i))], \ \ \forall z \in \mathbb{R}^k.$$

Note that

$$\mathbb{E}[G(\left(X_N^\omega \left(s_j \right) \right)_{j=1}^k, X_N^\omega(t_i), (X_N^\omega(r_j))_{j=1}^m, X_N^\omega(t_{i+1} -))] = \mathbb{E}\left[H_N \left(\left(X_N^\omega \left(s_j \right) \right)_{j=1}^k, X_N^\omega(t_i) \right) \right]. \tag{3.27}$$

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

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

and note that

$$\mathbb{E}[G((X^\omega \left(s_j \right))_{j=1}^k, X^\omega(t_i), (X^\omega(r_j))_{j=1}^m, X^\omega(t_{i+1} -))] = \mathbb{E}\left[H \left(\left(X^\omega \left(s_j \right) \right)_{j=1}^k, X^\omega(t_i) \right) \right]. \tag{3.28}$$

Using (3.26) and the Skorokhod representation theorem, we can assume that the random variables $\left(\left(X_N^\omega \left(s_j \right) \right)_{j=1}^k, X_N^\omega(t_i) \right)_{N \geq 1}$ and $\left(\left(X^\omega \left(s_j \right) \right)_{j=1}^k, X^\omega(t_i) \right)$ are defined on the same probability space and that the convergence holds almost surely. In particular, we can write

$$\left| \mathbb{E}\left[H_N \left(\left(X_N^\omega \left(s_j \right) \right)_{j=1}^k, X_N^\omega(t_i) \right) \right] - \mathbb{E}\left[H \left(\left(X^\omega \left(s_j \right) \right)_{j=1}^k, X^\omega(t_i) \right) \right] \right| \leq R_N^1 + R_N^2, \tag{3.29}$$

where

$$R_N^1 := \left| \mathbb{E}\left[H_N \left(\left(X_N^\omega \left(s_j \right) \right)_{j=1}^k, X_N^\omega(t_i) \right) \right] - \mathbb{E}\left[H_N((X^\omega \left(s_j \right))_{j=1}^k, X_N^\omega(t_i)) \right] \right|,$$
$$R_N^2 := \left| \mathbb{E}\left[H_N((X^\omega \left(s_j \right))_{j=1}^k, X_N^\omega(t_i)) \right] - \mathbb{E}\left[H \left(\left(X^\omega \left(s_j \right) \right)_{j=1}^k, X^\omega(t_i) \right) \right] \right|.$$

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

$$R_N^1 \leq \sum_{j=1}^k \mathbb{E}[|X_N^\omega \left(s_j \right) - X^\omega \left(s_j \right)|] \xrightarrow[N\to\infty]{} 0. \tag{3.30}$$

Moreover, since $X_N^\omega(t_i)$ converges to $X^\omega(t_i)$ almost surely, we conclude using Theorem 2.2 that, for any $z \in [0, 1]^k$, $H_N(z, X_N^\omega(t_i))$ converges to $H(z, X^\omega(t_i))$ almost surely. Therefore, using the dominated convergence theorem, we conclude that

$$R_N^2 \xrightarrow[N\to\infty]{} 0. \tag{3.31}$$

Plugging (3.30) and (3.31) into (3.29) and using (3.27) and (3.28) yields the result. □

4. The ASG and its relatives

In this section we formalize the definition of the quenched ASG, and we provide definitions for the k-ASG and the pLD-ASG.

4.1. Results related to Section 2.4: the quenched ASG

The aim of this section is to prove the following result.

Proposition 4.1. (Existence of the quenched ASG.) *Let $\omega \in \mathbb{D}^{\star}$. For any $n \in \mathbb{N}$ and $T > 0$, there is a branching–coalescing particle system $(\mathcal{G}_T^\omega(\beta))_{\beta \geq 0}$ starting at $\beta = 0-$ with n lines, that almost surely consists of finitely many lines at each time $\beta \in [0, T]$, and that satisfies the requirements (i), (ii), (iii'), (iv), and (v) of Definition 2.1.*

Proof. We will explicitly construct a branching–coalescing particle system $(\mathcal{G}_T^\omega(\beta))_{\beta \geq 0}$ with the desired properties. The main difficulty is that the environment ω may have infinitely many jumps on each compact interval. Fix $T > 0$ and $n \in \mathbb{N}$ (sampling size) and define

$$\Lambda_{\mathrm{mut}} := \left\{ \lambda_i^0, \lambda_i^1 \right\}_{i \geq 1}, \qquad \Lambda_{\mathrm{sel}} := \left\{ \lambda_i^{\triangle} \right\}_{i \geq 1}, \qquad \Lambda_{\mathrm{coal}} := \left\{ \lambda_{i,j}^{\blacktriangle} \right\}_{i,j \geq 1, i \neq j},$$

where λ_i^0, λ_i^1, $\lambda_i^{\triangle}$, and $\lambda_{i,j}^{\blacktriangle}$ are independent Poisson processes on $[0, T]$ with parameters $\theta \nu_0$, $\theta \nu_1$, σ, and 1, respectively. For $\beta \in [0, T]$, let $\tilde{\omega}(\beta) := \omega(T) - \omega((T - \beta)-)$ and $I_{\tilde{\omega}} := \{\beta \in [0, T] : \Delta \tilde{\omega}(\beta) > 0\}$; $I_{\tilde{\omega}}$ is the countable set of jump times of $\tilde{\omega}$. Let $\mathcal{U}_{\tilde{\omega}} := \{U_i(\beta)\}_{i \geq 1, \beta \in I_{\tilde{\omega}}}$ be an i.i.d. family of uniform random variables on $(0, 1)$. Assume, without loss of generality, that the arrival times of λ_i^0, λ_i^1, $\lambda_i^{\triangle}$, and $\lambda_{i,j}^{\blacktriangle}$, $i, j \in \mathbb{N}$, $i \neq j$, are countable, distinct from each other, and distinct from the jump times of $\tilde{\omega}$. Let I_{coal} (resp. I_{sel}) be the set of arrival times of Λ_{coal} (resp. Λ_{sel}).

We first construct a set $\mathcal{V}^\omega \subset \mathbb{N} \times [0, T]$ of *virtual lines*, representing the lines that would be part of the ASG if there were no coalescences. In particular, once a line enters this set, it will remain there. The set $\mathcal{V}^\omega$ is constructed on the basis of the set of potential branching times $I_{\mathrm{bran}} := I_{\tilde{\omega}} \cup I_{\mathrm{sel}}$ as follows. Consider the (countable) set

$$S_{\mathrm{bran}} := \{(\beta_1, \ldots, \beta_k) : k \in \mathbb{N}, \ 0 \leq \beta_1 < \cdots < \beta_k, \ \beta_i \in I_{\mathrm{bran}}, \ i \in [k]\},$$

and fix an injective function $i_\star : [n] \times S_{\mathrm{bran}} \to \mathbb{N} \setminus [n]$. The set $\mathcal{V}^\omega$ is determined as follows:

1. For any $i \in [n]$ (i.e. in the initial sample) and $\beta \in [0, T] : (i, \beta) \in \mathcal{V}^\omega$.

2. For any $(\beta_1, \ldots, \beta_k) \in S_{\mathrm{bran}}, j \in [n]$, and $\beta \in [\beta_k, T] : (i_\star(j, \beta_1, \ldots, \beta_k), \beta) \in \mathcal{V}^\omega$ if and only if

   - for any $\ell \in [k]$ with $\beta_\ell \in I_{\tilde{\omega}}$, $U_{i_\star(j, \beta_1, \ldots, \beta_{\ell-1})}(\beta_\ell) \leq \Delta \tilde{\omega}(\beta_\ell)$ (or $U_j(\beta_1) \leq \Delta \tilde{\omega}(\beta_1)$ if $\ell = 1$),
   - for any $\ell \in [k]$ such that $\beta_\ell \in I_{\mathrm{sel}}$, β_ℓ is a jump time of $\lambda_{i_\star(j, \beta_1, \ldots, \beta_{\ell-1})}^{\triangle}$ (or of $\lambda_j^{\triangle}$ if $\ell = 1$),

and these are all possible virtual lines; see Figure 7.

Let $V^\omega(\beta) := \{i \in \mathbb{N} : (i, \beta) \in \mathcal{V}^\omega\}$. According to Lemma 4.1 below, $V^\omega(\beta)$ is almost surely finite for all $\beta \in [0, T]$. Now, for $\beta \in I_{\mathrm{coal}}$, let (a_β, b_β) be the pair (i, j) such that β is an arrival time of $\lambda_{i,j}^{\blacktriangle}$. Since the Poisson processes $\lambda_{i,j}^{\blacktriangle}$, $i \neq j$, have distinct jump times, (a_β, b_β) is uniquely defined. Let

$$\tilde{I}_{\mathrm{coal}} := \{\beta \in I_{\mathrm{coal}} : a_\beta, b_\beta \in V^\omega(\beta)\} \quad \text{and} \quad \tilde{I}_{\mathrm{bran}} := \{\beta \in I_{\mathrm{bran}} : V^\omega(\beta-) \subsetneq V^\omega(\beta)\}.$$

Since $V^\omega(T)$ is independent of Λ_{coal} and almost surely finite, it follows that $\tilde{I}_{\mathrm{coal}}$ and $\tilde{I}_{\mathrm{bran}}$ are almost surely finite. Let $\beta_1 < \cdots < \beta_m$ be the elements of $\tilde{I}_{\mathrm{coal}} \cup \tilde{I}_{\mathrm{bran}}$ (set $\beta_0 := 0$
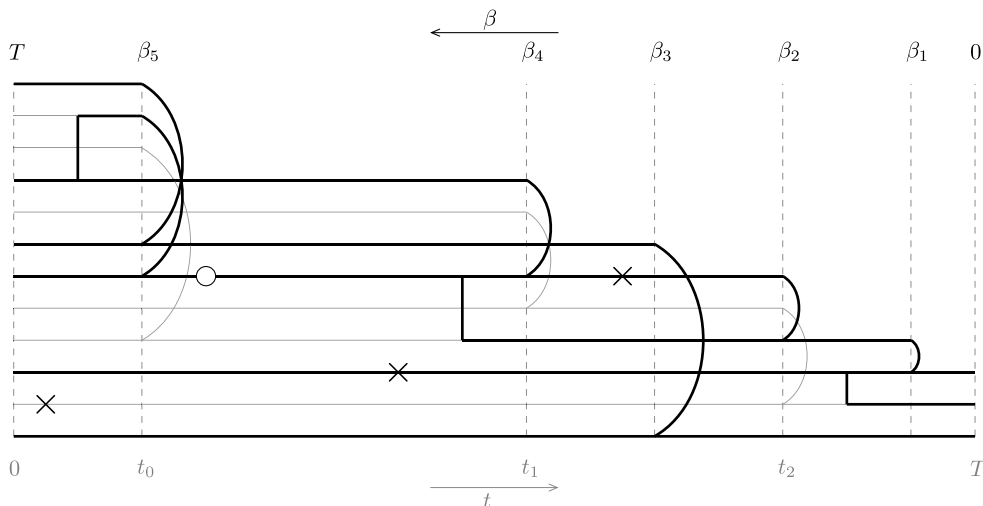
FIGURE 7. Illustration of the construction of the quenched ASG. The environment ω has jumps at forward times t_0, t_1, t_2; backward times $\beta_1, \ldots, \beta_5$ belong to the set of potential branching times $\tilde{I}_{\text{bran}}$. Virtual lines are depicted in grey or black; active lines are black. The ASG in $[0, T]$ consists of the set of active lines together with their connections and mutation marks.

and $\beta_{m+1} := T$ for convenience). We define $V_{\text{on}}^{\omega}(\beta) \subset V^{\omega}(\beta)$, the set of *active* lines at time β, as follows (see also Figure 7). For $\beta = 0$ we set $V_{\text{on}}^{\omega}(0) := V^{\omega}(0)$, and for $\beta \in (\beta_{\ell}, \beta_{\ell+1})$ we set $V_{\text{on}}^{\omega}(\beta) := V_{\text{on}}^{\omega}(\beta_{\ell})$. For $\beta = \beta_{\ell} \in \tilde{I}_{\text{coal}}$, we set $V_{\text{on}}^{\omega}(\beta_{\ell}) := V_{\text{on}}^{\omega}(\beta_{\ell} -) \setminus \{a_{\beta_{\ell}}\}$ if $\{a_{\beta_{\ell}}, b_{\beta_{\ell}}\} \subset V_{\text{on}}^{\omega}(\beta_{\ell} -)$, and $V_{\text{on}}^{\omega}(\beta_{\ell}) := V_{\text{on}}^{\omega}(\beta_{\ell} -)$ otherwise. Finally, for $\beta = \beta_{\ell} \in \tilde{I}_{\text{bran}}$, we set $V_{\text{on}}^{\omega}(\beta_{\ell}) := V_{\text{on}}^{\omega}(\beta_{\ell} -) \cup J_{\ell}$, where the set J_{ℓ} consists of the integers $i \in V^{\omega}(\beta_{\ell}) \setminus V^{\omega}(\beta_{\ell} -)$ such that $i = i_{\star}(j, \beta_{\ell})$ for some $j \in [n] \cap V_{\text{on}}^{\omega}(\beta_{\ell} -)$, or $i = i_{\star}(j, \hat{\beta}_1, \ldots, \hat{\beta}_k, \beta_{\ell})$ for some $(j, \hat{\beta}_1, \ldots, \hat{\beta}_k) \in i_{\star}^{-1}(V_{\text{on}}^{\omega}(\beta_{\ell} -) \setminus [n])$ with $\hat{\beta}_k < \beta_{\ell}$.

The ASG on $[0, T]$ is then the branching–coalescing system starting with n lines at levels in $[n]$, consisting at any time $\beta \in [0, T]$ of the lines in $V_{\text{on}}^{\omega}(\beta)$, and where the following hold:

(i) For $\beta \in I_{\text{bran}}$ such that $V_{\text{on}}^{\omega}(\beta -) \subsetneq V_{\text{on}}^{\omega}(\beta)$ and $i \in V_{\text{on}}^{\omega}(\beta) \setminus V_{\text{on}}^{\omega}(\beta -)$, either there is $(j, \hat{\beta}_1, \ldots, \hat{\beta}_k) \in [n] \times S_{\text{bran}}$ with $\hat{\beta}_k < \beta$ such that $i = i_{\star}(j, \hat{\beta}_1, \ldots, \hat{\beta}_k, \beta)$, or there is $j \in [n]$ such that $i = i_{\star}(j, \beta)$. In the first case, line $i_{\star}(j, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ branches at time β into $i_{\star}(j, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ (continuing line) and i (incoming line). In the second case, line j branches at time β into j (continuing line) and i (incoming line).

(ii) For $\beta \in I_{\text{coal}}$ such that $V_{\text{on}}^{\omega}(\beta) \subsetneq V_{\text{on}}^{\omega}(\beta -)$ and $i \in V_{\text{on}}^{\omega}(\beta -) \setminus V_{\text{on}}^{\omega}(\beta)$, $i = a_{\beta}$ and $b_{\beta} \in V_{\text{on}}^{\omega}(\beta)$. Thus, lines i and b_{β} merge into b_{β} at time β.

(iii) At each $\beta \in [0, T]$ that is an arrival time of λ_i^0 (resp. λ_i^1) for some $i \in V_{\text{on}}^{\omega}(\beta)$, we mark line i with a beneficial (resp. deleterious) mutation at time β.

Clearly the branching–coalescing particle system thus constructed satisfies the requirements (i), (ii), (iii'), (iv), and (v) of Definition 2.1. This ends the proof. $\square$

It remains to prove the following lemma.

Lemma 4.1. *The set $V^{\omega}(\beta)$ is almost surely finite for any $\beta \in [0, T]$.*

Proof. We keep using the notation introduced in the proof of Proposition 4.1. For $\delta > 0$, we consider the environment ω^δ defined via (3.1). We couple the sets of virtual lines $\mathcal{V}^\omega$ and $\mathcal{V}^{\omega^\delta}$ associated to ω and ω^δ, respectively, by using the same random sets Λ_{sel} and $\mathcal{U}_{\tilde{\omega}}$ (note that for $\beta \in I_{\tilde{\omega}}$ with $\Delta\tilde{\omega}(\beta) < \delta$, $\Delta\tilde{\omega}^\delta(\beta) = 0 < U_i(\beta)$). Let $N_T^\omega(\beta) := |V^\omega(\beta)|$ and $N_T^{\omega^\delta}(\beta) := |V^{\omega^\delta}(\beta)|$, $\beta \in [0, T]$. Since $\beta \mapsto N_T^\omega(\beta)$ is non-decreasing, it is enough to prove that $N_T^\omega(T) < \infty$ almost surely. From the construction of the set of virtual lines, it follows that $N_T^{\omega^\delta}(\beta)$ increases almost surely to $N_T^\omega(\beta)$ as $\delta \to 0$. By the monotone convergence theorem, for all $\beta \in [0, T]$ we get

$$\lim_{\delta \to 0} \mathbb{E}\left[N_T^{\omega^\delta}(\beta) \mid N_T^{\omega^\delta}(0) = n\right] = \mathbb{E}\left[N_T^\omega(\beta) \mid N_T^\omega(0) = n\right]. \tag{4.1}$$

Recall that $\tilde{\omega}^\delta(\beta) := \omega^\delta(T) - \omega^\delta((T - \beta)-)$, $\beta \in [0, T]$, and that $\tilde{\omega}^\delta$ has finitely many jumps in $[0, T]$. Let $T_1 < \cdots < T_N$ be the jump times of $\tilde{\omega}^\delta$. The process $\left(N_T^{\omega^\delta}(\beta)\right)_{\beta \in [0,T]}$ has the following transitions:

1. On $(T_i, T_{i+1}) : N_T^{\omega^\delta}$ jumps from k to $k + 1$ at rate $k\sigma$.

2. At time T_i : $N_T^{\omega^\delta}$ jumps from k to $k + B_k(\Delta\tilde{\omega}^\delta(T_i))$, where $B_k(\Delta\tilde{\omega}^\delta(T_i)) \sim$ Bin$(k, \Delta\tilde{\omega}^\delta(T_i))$.

Note that for each T_i we have $\Delta\tilde{\omega}^\delta(T_i) = \Delta\tilde{\omega}(T_i)$. This yields in particular

$$\mathbb{E}[N_T^{\omega^\delta}(T_i) \mid N_T^{\omega^\delta}(T_i-)] = (1 + \Delta\tilde{\omega}(T_i))N_T^{\omega^\delta}(T_i-). \tag{4.2}$$

Successively using Lemma 4.2 (see below) and (4.2), we get

$$\mathbb{E}\left[N_T^{\omega^\delta}(T) \mid N_T^{\omega^\delta}(0) = n\right] \leq n e^{\sigma T} \prod_{\beta \in [0,T]: \Delta\tilde{\omega}(\beta) \geq \delta} (1 + \Delta\tilde{\omega}(\beta)).$$

In particular, we have

$$\mathbb{E}[N_T^{\omega^\delta}(T) \mid N_T^{\omega^\delta}(0) = n] \leq n e^{\sigma T} \prod_{\beta \in I_{\tilde{\omega}} \cap [0,T]} (1 + \Delta\tilde{\omega}(\beta)) < \infty. \tag{4.3}$$

Letting δ go to 0 in (4.3) and using (4.1), we get

$$\mathbb{E}[N_T^\omega(T) \mid N_T^\omega(0) = n] \leq n e^{\sigma T} \prod_{\beta \in I_{\tilde{\omega}} \cap [0,T]} (1 + \Delta\tilde{\omega}(\beta)) < \infty.$$

This concludes the proof. □

Lemma 4.2. *Let $0 \leq \beta_1 < \beta_2 \leq T$ be such that $\tilde{\omega}^\delta$ has no jump times on $(\beta_1, \beta_2]$. Then we have*

$$\mathbb{E}[N_T^{\omega^\delta}(\beta_2) \mid N_T^{\omega^\delta}(\beta_1)] \leq e^{\sigma(\beta_2 - \beta_1)} N_T^{\omega^\delta}(\beta_1).$$

Proof. Since $\tilde{\omega}^\delta$ has no jump times on $(\beta_1, \beta_2]$, on this interval $N_T^{\omega^\delta}$ is the Markov chain on $\mathbb{N}$ with generator $\mathcal{G}_N f(n) = \sigma n(f(n + 1) - f(n))$. Let $f_M(n) := n \wedge M$. Note that, for any

$M, n \geq 1$, we have $\mathcal{G}_N f_M(n) \leq \sigma f_M(n)$. Applying Dynkin's formula to $N_T^{\omega^\delta}$ on $(\beta_1, \beta_2]$ with the function f_M, we obtain

$$
\begin{aligned}
\mathbb{E}\left[f_M(N_T^{\omega^\delta}(\beta_2)) \mid N_T^{\omega^\delta}(\beta_1)\right] &= f_M(N_T^{\omega^\delta}(\beta_1)) + \mathbb{E}\left[\int_{\beta_1}^{\beta_2} \mathcal{G}_N f_M(N_T^{\omega^\delta}(\beta)) \mathrm{d}\beta \mid N_T^{\omega^\delta}(\beta_1)\right] \\
&\leq f_M(N_T^{\omega^\delta}(\beta_1)) + \sigma \mathbb{E}\left[\int_{\beta_1}^{\beta_2} f_M(N_T^{\omega^\delta}(\beta)) \mathrm{d}\beta \mid N_T^{\omega^\delta}(\beta_1)\right] \\
&= f_M(N_T^{\omega^\delta}(\beta_1)) + \sigma \int_{\beta_1}^{\beta_2} \mathbb{E}\left[f_M(N_T^{\omega^\delta}(\beta)) \mid N_T^{\omega^\delta}(\beta_1)\right] \mathrm{d}\beta.
\end{aligned}
$$

Thus, Gronwall's lemma yields $\mathbb{E}[f_M(N_T^{\omega^\delta}(\beta_2)) \mid N_T^{\omega^\delta}(\beta_1)] \leq e^{\sigma(\beta_2 - \beta_1)} f_M(N_T^{\omega^\delta}(\beta_1))$. The result follows from letting $M \to \infty$ and using the monotone convergence theorem. $\qquad\square$

4.2. Definitions related to Section 2.5: the killed ASG

The k-ASG as a branching–coalescing system of particles is defined as follows (see Figure 4).

Definition 4.1. *(The annealed/quenched k-ASG.)* The *annealed* k-ASG with parameters $\sigma, \theta, \nu_0, \nu_1$, and environment driven by a pure-jump subordinator with Lévy measure μ, of a sample of size n is the branching–coalescing particle system $\bar{\mathcal{G}} := (\bar{\mathcal{G}}(\beta))_{\beta \geq 0}$ starting with n lines and with the following dynamic:

  (i) Each line splits into two lines, an incoming line and a continuing line, at rate σ.

  (ii) Every given pair of lines coalesces into a single line at rate 2.

  (iii) Every group of k lines is subject to a simultaneous branching at rate $\sigma_{m,k}$ (defined in Equation (2.5)), 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.

  (iv) Each line is killed at rate $\theta \nu_1$.

  (v) Each line sends the process to the cemetery state $\dagger$ at rate $\theta \nu_0$.

Let $\omega \in \mathbb{D}^\star$. The *quenched k-ASG* with parameters $\sigma, \theta, \nu_0, \nu_1$, and environment ω, of a sample of size n at time T is the branching–coalescing particle system $\bar{\mathcal{G}}_T^\omega := (\bar{\mathcal{G}}_T^\omega(\beta))_{\beta \geq 0}$ starting at $\beta = 0-$ with n lines and evolving according to (i), (ii), (iv), and (v) of the previous definition, with (iii) replaced by the following:

  (iii$'$) If at time β we have $\Delta\omega(T - \beta) > 0$, then any line splits into two lines, an incoming line and a continuing line, with probability $\Delta\omega(T - \beta)$, independently from the other lines.

Remark 4.1. The branching–coalescing system underlying the quenched k-ASG is well-defined, because it can be constructed on the basis of the quenched ASG.

4.3. Definitions related to Section 2.6: the pruned lookdown ASG

In this section, we give a detailed construction of the pLD-ASG, which incorporates the effect of the environment.
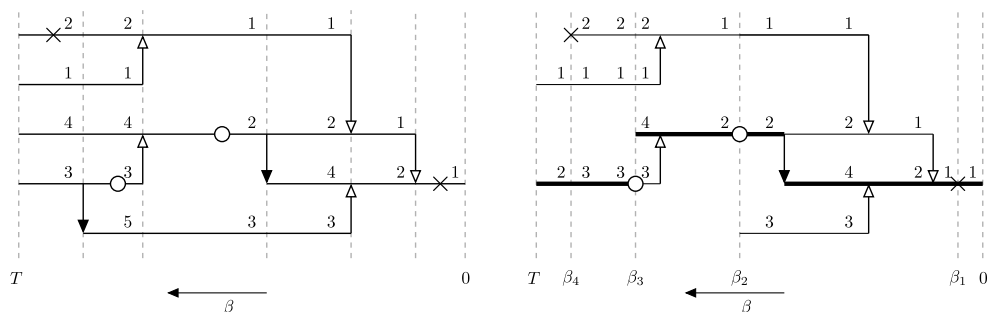
FIGURE 8. LD-ASG (left) and its pLD-ASG (right). Backward time $\beta \in [0, T]$ runs from right to left. In the LD-ASG, levels remain constant between the dashed lines; in particular, they are not affected by mutation events. In the pLD-ASG, lines are pruned at mutation events, where an additional updating of the levels takes place. The bold line in the pLD-ASG represents the immune line.

First, we construct the (annealed/quenched) *lookdown ASG* (LD-ASG). The latter is the ASG equipped with a numbering of its lines encoding the hierarchy given by the pecking order. This is done as follows. Consider a realization of the (annealed/quenched) ASG in $[0, T]$ starting with one line, which is assigned level 1. When the line at level i coalesces with the line at level $j > i$, the resulting line is assigned level i; the level of each line having 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. $i_k + k$); 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 Figure 8 (left panel) for an illustration. The pLD-ASG is obtained via an appropriate pruning of the lines of the LD-ASG. Before describing the pruning procedure, we identify a special line in the LD-ASG: *the immune line*. The immune line at time β is the line in the ASG present at time β that is the ancestor of the starting line if all the lines at time β are assigned the unfit type. In the absence of mutations, the immune line changes only if it is involved in a coalescence or branching event. If it is involved in a coalescence event, the merged line is the new immune line. If it is involved in a branching event, the continuing line is the new immune line.

In the presence of mutations, the pLD-ASG is constructed simultaneously with the immune line as follows. Let $\beta_1 < \cdots < \beta_m$ be the times at which mutations occur in the LD-ASG in $[0, T]$. In the time interval $[0, \beta_1)$, the pLD-ASG coincides with the LD-ASG and the immune line evolves as before. Now, assume that we have constructed the pLD-ASG together with its immune line up to time β_i-, where the pLD-ASG contains n lines and the immune line has level $k_0 \in [n]$. The pLD-ASG is extended up to time β_i according to the following rules:

(i) If, at time β_i, a line with level $k \neq k_0$ at β_i- is hit by a deleterious mutation, we stop tracing back this line; all the other lines are extended up to time β_i; all the lines with level $j > k$ at time β_i- decrease their level by 1, and the others keep their levels unchanged; the immune line continues on the same line (possibly with a different level).

(ii) If, at time β_i-, the line with level k_0 at β_i- is hit by a deleterious mutation, we extend all the lines up to time β_i; the immune line gets level n, but remains on the same line; all lines having a level $j > k_0$ at time β_i- decrease their level by 1, and the others keep their levels unchanged.

(iii) If, at time β_i, a line with level k is hit by a beneficial mutation, we stop tracing back all the lines with level $j > k$; the remaining lines are extended up to time β_i, keeping their levels; the line hit by the mutation becomes the immune line.

In $[\beta_i, \beta_{i+1})$, $i \in [m-1]$, and $[\beta_m, T]$, the pLD-ASG evolves as the LD-ASG, and the immune line as in the case without mutations. The next result states the main feature of the pLD-ASG.

Lemma 4.3. *If we assign types at (backward) time T in the pLD-ASG, the true ancestor of the single line at (backward) time 0 is the line of type 0 with smallest level, or, if all lines have type 1, it is the immune line.*

Proof. The proof is analogous to the proof of [31, Theorem 4], which covers the null environment case. □

5. Annealed results

5.1. Annealed results related to Section 2.5

We start this section by proving the first part of Theorem 2.3, i.e. Equations (2.7) and (2.8).

Proof of Theorem 2.3 (Part I: reinforced and annealed moment duality). Define the function $H : [0, 1] \times \mathbb{N}_0^\dagger \times [0, \infty)$ via $H(x, n, j) := (1-x)^n f(j)$. Let $(P_t)_{t \geq 0}$ and $(Q_t)_{t \geq 0}$ denote the semigroups of (X, J) and (R, J), respectively, i.e.

$$P_t g(x, j) = \mathbb{E}[g(X(t), J(t) + j) \mid X(0) = x],$$

$$Q_t h(n, j) = \mathbb{E}[h(R(t), J(t) + j) \mid R(0) = n].$$

Let $(\hat{R}, \hat{J})$ be a copy of (R, J), which is independent of (X, J). A straightforward calculation shows that

$$P_t(Q_s H)(x, n, j) = \mathbb{E}[(1 - X(t))^{\hat{R}(s)} f(J(t) + \hat{J}(s) + j) \mid X(0) = x, \ \hat{R}(0) = n]$$

$$= Q_s(P_t H)(x, n, j). \tag{5.1}$$

Let G and $G_\star$ be the infinitesimal generators of (X, J) and (R, J), respectively. Clearly, for any $x \in [0, 1]$, the function $(n, j) \mapsto P_t H(\cdot, n, \cdot)(x, j)$ belongs to the domain of $G_\star$. Hence, Equation (5.1) yields

$$P_t G_\star H(x, n, j) = G_\star P_t H(x, n, j). \tag{5.2}$$

We claim that

$$G H(\cdot, n, \cdot)(x, j) = G_\star H(x, \cdot, \cdot)(n, j). \tag{Claim 4}$$

Assume that Claim 4 holds. Define the functions $u(t, x, n, j) := P_t H(\cdot, n, \cdot)(x, j)$ and $v(t, x, n, j) := Q_t H(x, \cdot, \cdot)(n, j)$. The Kolmogorov forward equation for Q yields

$$\frac{\mathrm{d}}{\mathrm{d}t} v(t, x, n, j) = G_\star v(x, \cdot, \cdot)(n, j). \tag{5.3}$$

Moreover, using the Kolmogorov forward equation for P, Claim 4, and (5.2), we get

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t, x, n, j) = P_t G H(\cdot, n, \cdot)(x, j) = P_t G_\star H(x, \cdot, \cdot)(n, j) = G_\star u(x, \cdot, \cdot)(n, j).$$

Hence, u and v satisfy Equation (5.3). Since $u(0, x, n, j) = (1 - x)^n f(j) = v(0, x, n, j)$, Equation (2.7) follows from the uniqueness of the initial value problem associated with $G_\star$ (see [14, Theorem 1.3]). Equation (2.8) is obtained using $f \equiv 1$ in Equation (2.7). It remains to prove Claim 4. Note first that

$$GH(\cdot, n, \cdot)(x, j) = (1 - x)^n \int_{(0,1]} \left[(1 - xz)^n f(j + z) - f(j) \right] \mu(dz)$$
$$+ \left[n(n-1)x(1-x)^{n-1} - (\sigma x(1-x) + \theta \nu_0(1-x) - \theta \nu_1 x) \, n(1-x)^{n-1} \right] f(j). \tag{5.4}$$

In addition,

$$G_\star H(x, \cdot, \cdot)(n, j) = n((n-1) + \theta \nu_1)[(1-x)^{n-1} - (1-x)^n] f(j)$$
$$+ \sigma n[(1-x)^{n+1} - (1-x)^n] f(j) - n\theta \nu_0 (1-x)^n f(j)$$
$$+ \sum_{k=0}^{n} \binom{n}{k} \int_{(0,1)} y^k (1-y)^{n-k} [(1-x)^{n+k} f(j+y) - (1-x)^n f(j)] \mu(dy)$$
$$= \left[n(n-1)x(1-x)^{n-1} - (\sigma x(1-x) + \theta \nu_0(1-x) - \theta \nu_1 x) \, n(1-x)^{n-1} \right] f(j)$$
$$+ (1-x)^n \sum_{k=0}^{n} \binom{n}{k} \int_{(0,1)} y^k (1-y)^{n-k} [(1-x)^k f(j+y) - f(j)] \mu(dy). \tag{5.5}$$

Moreover, using Fubini's theorem, we obtain

$$\sum_{k=0}^{n} \binom{n}{k} \int_{(0,1)} y^k (1-y)^{n-k} [(1-x)^k f(j+y) - f(j)] \mu(dy)$$
$$= \int_{(0,1)} \left[(1 - xy)^n f(j+y) - f(j) \right] \mu(dy).$$

Hence, Claim 4 follows after comparing (5.5) with (5.4). □

We now prove Theorem 2.4(1), which characterizes the asymptotic type frequency in the annealed setting.

Proof of Theorem 2.4(1) *(asymptotic type frequency).* We first show that $X(t)$ has a limit in distribution as $t \to \infty$. Since $\theta > 0$ and $\nu_0 \in (0, 1)$, Equation (2.8) in Theorem 2.3 implies that, for any $x \in [0, 1]$, the limit of $\mathbb{E}[(1 - X(t))^n \mid X(0) = x]$ as $t \to \infty$ exists and satisfies

$$\lim_{t \to \infty} \mathbb{E}[(1 - X(t))^n | X(0) = x] = \pi_n, \quad n \in \mathbb{N}_0, \tag{5.6}$$

where π_n is defined in (2.10). Recall that probability measures on [0, 1] are completely determined by their positive integer moments and that convergence of positive integer moments implies convergence in distribution. Therefore, Equation (5.6) implies that there is $\eta_X \in \mathcal{M}_1([0, 1])$ such that, for any $x \in [0, 1]$, conditionally on $\{X(0) = x\}$, the law of $X(t)$ converges in distribution to η_X as $t \to \infty$ and

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

Using dominated convergence, the convergence of the law of $X(t)$ towards η_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 η_X.

Finally, a first step decomposition for the probability of absorption in 0 of R yields

$$\left[n(\sigma + \theta + n - 1) + \sum_{k=1}^{n} \binom{n}{k} \sigma_{n,k} \right] \pi_n = n\sigma \pi_{n+1} + n(\theta \nu_1 + n - 1)\pi_{n-1} + \sum_{k=1}^{n} \binom{n}{k} \sigma_{n,k} \pi_{n+k}.$$

Dividing both sides in the previous identity by n and rearranging terms yields Equation (2.12). $\square$

5.2. Annealed results related to Section 2.6

In this section we prove Theorem 2.5(1) and Corollary 2.1. Before that we prove the following lemma relating the ancestral type distribution at time T to the number $L(T)$ of lines in the pLD-ASG at time T.

Lemma 5.1. *For all $T \geq 0$ and $x \in [0, 1]$, we have*

$$h_T(x) = 1 - \mathbb{E}[(1-x)^{L(T)} \mid L(0) = 1]. \tag{5.7}$$

Proof. Since types are assigned to the $L(T)$ lines present in the pLD-ASG at (backward) time T according to independent Bernoulli random variables with parameter x, the result follows from Lemma 4.3. $\square$

The next result is crucial for describing the asymptotic behavior of $h_T(x)$ as $T \to \infty$.

Lemma 5.2. (Positive recurrence.) *The process L is positive recurrent.*

Proof. Since L is irreducible, it is enough to prove that the state 1 is positive recurrent. This holds if $\theta \nu_0 > 0$, because in this case the hitting time of 1 is bounded above by an exponential random variable with parameter $\theta \nu_0$. Now, assume that $\theta = 0$ (the case $\theta \nu_0 = 0$ and $\theta \nu_1 > 0$ can easily be reduced to this case). We proceed in a way similar to the proof of [16, Lemma 2.3]. Define the function $f : \mathbb{N} \to \mathbb{R}_+$ via

$$f(n) := \sum_{i=1}^{n-1} \frac{1}{i} \ln \left(1 + \frac{1}{i} \right),$$

with the convention that an empty sum equals 0. Note that f is bounded. Note also that, for $n > 1$,

$$n(n-1)(f(n-1) - f(n)) = -n \ln \left(1 + \frac{1}{n-1} \right) \leq -1.$$

This follows from using $x = 1/n$ in the inequality $e^x < 1/(1-x)$, which holds for $x < 1$. For any $\varepsilon > 0$, set $n_0(\varepsilon) := \lfloor 1/\varepsilon \rfloor + 1$. Note that for $n > n_0(\varepsilon)$,

$$n(f(n+i) - f(n)) = n \sum_{j=n}^{n+i-1} \frac{1}{j} \ln \left(1 + \frac{1}{j} \right) \leq n \ln \left(1 + \frac{1}{n} \right) i\varepsilon \leq i\varepsilon.$$

Hence, for $n > n_0(\varepsilon)$,

$$G_L f(n) \le -1 + \frac{\varepsilon}{n} \sum_{i=1}^{n} \binom{n}{i} \sigma_{n,i}\, i + \sigma \varepsilon = -1 + \varepsilon \int_{(0,1)} \sum_{i=1}^{n} \binom{n-1}{i-1} y^i (1-y)^{n-i} \mu(dy) + \sigma \varepsilon$$

$$= -1 + \varepsilon \int_{(0,1)} y \mu(dy) + \sigma \varepsilon,$$

where $\sigma_{n,i}$ is defined in (2.5). Set $m_0 := n_0(\varepsilon_\star)$, where $\varepsilon_\star := 1 / \left(2 \int_{(0,1)} y \mu(dy) + 2\sigma \right)$ (and we set $m_0 := 1$ in the particular case $\mu = 0$ and $\sigma = 0$). In particular, for $n > m_0$, we have $G_L f(n) \le -1/2$.

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

$$\mathbb{E}\left[f(L(T_{m_0} \wedge k)) \mid L(0) = n \right] = f(n) + \mathbb{E}\left[\int_0^{T_{m_0} \wedge k} G_L f(L(\beta)) d\beta \mid L(0) = n \right].$$

Therefore, for $n > m_0$, we have

$$0 \le \mathbb{E}\left[f(L(T_{m_0} \wedge k)) \mid L(0) = n \right] \le f(n) - \frac{1}{2} \mathbb{E}[T_{m_0} \wedge k \mid L(0) = n].$$

In particular, we have $\mathbb{E}[T_{m_0} \wedge k \mid L(0) = n] \le 2f(n)$. Thus, letting $k \to \infty$ in this inequality yields $\mathbb{E}[T_{m_0} \mid L(0) = n] \le 2f(n) < \infty$. Since L is irreducible, the result follows by standard arguments. $\qquad\square$

The first part of the proof of Theorem 2.5(1) builds on the previous two lemmas. The system of equations (2.18) characterizing the tail probabilities $\mathbb{P}(L(\infty) > n)$ is obtained via Siegmund duality. More precisely, consider the continuous-time Markov chain $D := (D(\beta))_{\beta \ge 0}$ with values in $\mathbb{N}^\dagger := \mathbb{N} \cup \{\dagger\}$ with rates

$$q_D(i,j) := \begin{cases} (i-1)(\sigma + \sigma_{i-1,1}) & \text{if } j = i-1,\, i > 1, \\ (i-1)\theta\nu_1 + i(i-1) & \text{if } j = i+1,\, i > 1, \\ \gamma_{i,j} - \gamma_{i,j-1} & \text{if } 1 \le j < i,\, i > 2, \\ (i-1)\theta\nu_0 & \text{if } j = \dagger,\, i > 1, \end{cases}$$

where $\dagger$ is a cemetery point, and where $\gamma_{i,j} := \sum_{k=i-j}^{j} \binom{j}{k} \sigma_{j,k}$ if $1 \le j < i \le 2j$ and $\gamma_{i,j} := 0$ otherwise (see Theorem 2.5). The states 1 and $\dagger$ are absorbing for D. The next result relates L and D via duality.

Lemma 5.3. (Siegmund duality.) *The processes L and D are Siegmund dual; i.e.*

$$\mathbb{P}\left(L(\beta) \ge d \mid L(0) = \ell \right) = \mathbb{P}\left(\ell \ge D(\beta) \mid D(0) = d \right) \qquad \text{for all } \ell, d \in \mathbb{N},\ t \ge 0.$$

Proof. Define $H : \mathbb{N} \times \mathbb{N} \cup \{\dagger\} \to \{0, 1\}$ via $H(\ell, d) := 1_{\{\ell \ge d\}}$ 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 [23, Proposition 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 (2.16), we have

$$G_L H(\cdot, d)(\ell)$$

$$= \sigma\ell \, 1_{\{\ell+1=d\}} - (\ell-1)(\ell+\theta\nu_1) \, 1_{\{\ell=d\}} - \theta\nu_0 \sum_{j=1}^{\ell-1} 1_{\{j<d\leq\ell\}} + \sum_{k=1}^{\ell} \binom{\ell}{k} \sigma_{\ell,k} \, 1_{\{\ell<d\leq\ell+k\}}$$

$$= \sigma\ell \, 1_{\{\ell+1=d\}} - (\ell-1)(\ell+\theta\nu_1) \, 1_{\{\ell=d\}} - \theta\nu_0(d-1)1_{\{d\leq\ell\}} + \gamma_{d,\ell} \, 1_{\{\ell<d\}}. \quad (5.8)$$

Similarly, we have

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

$$+ \sum_{j=1}^{d-1} \left(\gamma_{d,j} - \gamma_{d,j-1}\right) \, 1_{\{j\leq\ell<d\}}. \quad (5.9)$$

Summation by parts yields $\sum_{j=1}^{d-1} \left(\gamma_{d,j} - \gamma_{d,j-1}\right) \, 1_{\{j\leq\ell<d\}} = \gamma_{d,\ell} 1_{\{\ell<d\}}$. Thus, the result follows from comparing (5.9) with (5.8). □

Now we have all the ingredients to prove Theorem 2.5(1).

Proof of Theorem 2.5(1) *(ancestral type distribution).* Since L is positive recurrent, $L(T)$ converges in distribution as $T \to \infty$ towards the stationary distribution η_L. In particular, we infer from Equation (5.7) that $h(x) := \lim_{T\to\infty} h_T(x)$ exists and satisfies

$$h(x) = 1 - \mathbb{E}[(1-x)^{L(\infty)}] = 1 - \sum_{\ell=1}^{\infty} \mathbb{P}(L(\infty) = \ell)(1-x)^{\ell}$$

$$= \sum_{\ell=0}^{\infty} \mathbb{P}(L(\infty) > \ell)(1-x)^{\ell} - (1-x) \sum_{\ell=1}^{\infty} \mathbb{P}(L(\infty) > \ell-1)(1-x)^{\ell-1},$$

and Equation (2.17) follows. It remains to prove (2.18). From Lemma 5.3 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, for $n > 1$,

$$\left((n-1)(\sigma + \theta + n) + \gamma_{n,n-1}\right) d_n$$

$$= (n-1)\sigma d_{n-1} + (n-1)(\theta\nu_1 + n)d_{n+1} + \sum_{j=1}^{n-1} (\gamma_{n,j} - \gamma_{n,j-1})d_j. \quad (5.10)$$

Using summation by parts and rearranging terms in (5.10) yields

$$(\sigma + \theta + n)d_n = \sigma d_{n-1} + (\theta\nu_1 + n)d_{n+1} + \frac{1}{n-1} \sum_{j=1}^{n-1} \gamma_{n,j}(d_j - d_{j+1}), \quad n > 1. \quad (5.11)$$

The result follows. □

Proof of Corollary 2.1. Since $\theta = 0$, the line-counting processes R and L have the same distribution. Hence, combining Lemma 5.1 and (2.8) (from Theorem 2.3) applied to $n = 1$, we obtain

$$h_T(x) = \mathbb{E}[X(T) \mid X(0) = x], \tag{5.12}$$

which proves the first part of the statement. Moreover, for $\theta = 0$, X is a bounded submartingale, and hence $X(T)$ almost surely has a limit as $T \to \infty$, which we denote by $X(\infty)$. Letting $T \to \infty$ in the identity (5.12) yields

$$h(x) = \mathbb{E}[X(\infty) \mid X(0) = x]. \tag{5.13}$$

Moreover, using (2.8) (from Theorem 2.3) 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 (5.13) into the previous identity yields the desired result. $\qquad\square$

Proof of Proposition 2.2 Using Equation (2.18) in Theorem 2.5 for the two models, we obtain, for $n \in \mathbb{N}$,

$$(n+1)\rho_{n+1}^{\mathrm{sel}} = \sigma_\mu \rho_n^{\mathrm{sel}}, \quad \text{and} \quad (n+1)\rho_{n+1}^{\mathrm{env}} = \frac{1}{n}\sum_{j=1}^{n}\gamma_{n+1,j}\,\rho_j^{\mathrm{env}}.$$

Separately multiplying these equations by z^n, $z \in [0, 1]$, and summing over $n \in \mathbb{N}$, one obtains

$$(p^{\mathrm{sel}})'(z) = \rho_1^{\mathrm{sel}} + \sigma_\mu\, p^{\mathrm{sel}}(z), \quad \text{and} \quad (p^{\mathrm{env}})'(z) = \rho_1^{\mathrm{env}} + \sum_{j=1}^{\infty}\rho_j^{\mathrm{env}}g_j(z),$$

with $g_j(z) := \sum_{n=j}^{2j-1} \gamma_{n+1,j}\frac{z^n}{n}$. Solving the ordinary differential equation for p^{sel} via variation of constants, and using that $p^{\mathrm{sel}}(0)$ and $p^{\mathrm{sel}}(1) = 1$, we obtain the desired formulas for ρ_1^{sel} and p^{sel} (see also [13, Theorem 6.1]). Now, using the definitions of the coefficients $\gamma_{n+1,j}$ (defined below Equation (2.18)) and $\sigma_{m,k}$ (see (2.5)) followed by a straightforward calculation, one obtains

$$g_j(z) = \sum_{k=1}^{j}\binom{j}{k}\sigma_{j,k}\int_0^z \frac{u^{j-1} - u^{k+j-1}}{1-u}\,\mathrm{d}u = \int_0^z \mathrm{d}u\,\frac{u^{j-1}}{1-u}\int_{(0,1)}\mu(\mathrm{d}y)\,(1 - (1 - y(1-u))^j).$$

Since $(1 - h)^j \geq 1 - jh$ for $h \in (0, 1)$, we infer that $g_j(z) \leq \sigma_\mu z^j$, with equality only if $z = 0$ or $j = 1$. We conclude that

$$(p^{\mathrm{env}})'(z) < \rho_1^{\mathrm{env}} + \sigma_\mu\, p^{\mathrm{env}}(z), \quad z \in (0, 1].$$

Letting $f(z) := \rho_1^{\mathrm{env}} + \sigma_\mu\, p^{\mathrm{env}}(z)$, we then have $f'(z)/f(z) \leq \sigma_\mu$, so, after integration, $\log(f(z)/f(0)) \leq \sigma_\mu z$. Since $p^{\mathrm{env}}(0) = 0$, this yields

$$p^{\mathrm{env}}(z) \leq \rho_1^{\mathrm{env}}\left(\frac{e^{\sigma_\mu z} - 1}{\sigma_\mu}\right) = \frac{\rho_1^{\mathrm{env}}}{\rho_1^{\mathrm{sel}}}\, p^{\mathrm{sel}}(z).$$

Moreover, since $p^{\mathrm{env}}(1) = p^{\mathrm{sel}}(1) = 1$, we conclude that $\rho_1^{\mathrm{env}} \geq \rho_1^{\mathrm{sel}}$. Assume now that $\rho_1^{\mathrm{env}} = \rho_1^{\mathrm{sel}}$. It follows that $p^{\mathrm{env}}(z) \leq p^{\mathrm{sel}}(z)$ for $z \in [0, 1]$. Hence,

$$1 = \int_0^1 (p^{\mathrm{env}})'(z)\mathrm{d}z < \int_0^1 (\rho_1^{\mathrm{env}} + \sigma_\mu \, p^{\mathrm{env}}(z))\mathrm{d}z \leq \int_0^1 (\rho_1^{\mathrm{sel}} + \sigma_\mu \, p^{\mathrm{sel}}(z))\mathrm{d}z = \int_0^1 (p^{\mathrm{sel}})'(z)\mathrm{d}z = 1,$$

which is a contradiction. Thus, $\rho_1^{\mathrm{env}} > \rho_1^{\mathrm{sel}}$. In particular, for $z \neq 0$ sufficiently small, $p^{\mathrm{env}}(z) > p^{\mathrm{sel}}(z)$. Hence, the last statement follows from the fact that $h^{\mathrm{env}}(z) = 1 - p^{\mathrm{env}}(1 - z)$ and $h^{\mathrm{sel}}(z) = 1 - p^{\mathrm{sel}}(1 - z)$. □

6. Quenched results

6.1. Quenched results related to Section 2.5

In this section we prove the quenched parts of the results stated in Section 2.5. We start with the proof of the second part of Theorem 2.3, which establishes the quenched moment duality (2.9) for almost every environment ω.

Proof of Theorem 2.3 (Part II: quenched moment duality). Since both sides of (2.9) are right-continuous in T, it is sufficient to prove that, for any bounded measurable function $g : \mathbb{D}_T^\star \to \mathbb{R}$,

$$\mathbb{E}[(1 - X^J(T))^n g((J_s)_{s \in [0,T]}) | X^J(0) = x] = \mathbb{E}[(1 - x)^{R_T^J(T-)} g((J_s)_{s \in [0,T]}) | R_T^J(0-) = n]. \quad (6.1)$$

Let $\mathcal{H} := \{g : \mathbb{D}_T^\star \to \mathbb{R} : \text{ such that holds}\}$. Thanks to the annealed moment duality, Equation (2.8), every constant function belongs to $\mathcal{H}$. Moreover, $\mathcal{H}$ is closed under increasing limits of non-negative bounded functions in $\mathcal{H}$. We claim that (6.1) holds for functions of the form $g(\omega) = g_1(\omega(t_1)) \cdots g_k(\omega(t_k))$, with $0 < t_1 < \cdots < t_k < T$ and $g_i \in \mathcal{C}^2([0, \infty))$ with compact support. If the claim is true, then thanks to the monotone class theorem, $\mathcal{H}$ will contain any measurable function g, which will then complete the proof.

We prove the claim by induction on k. For $k = 1$, we need to prove that for $t_1 \in (0, T)$,

$$\mathbb{E}[(1 - X^J(T))^n g_1(J(t_1)) | X^J(0) = x] = \mathbb{E}[(1 - x)^{R_T^J(T-)} g_1(J(t_1)) | R_T^J(0-) = n]. \quad (6.2)$$

Note first that, using the Markov property for X^J in $[0, t_1]$, Equation (2.8), and the fact that t_1 and T are almost surely continuity times for J, we obtain

$$\mathbb{E}\left[(1 - X^J(T))^n g_1(J(t_1)) \,|\, X^J(0) = x\right]$$
$$= \mathbb{E}\left[g_1(J(t_1))\hat{\mathbb{E}}\left[(1 - \hat{X}^{\hat{J}}(T - t_1))^n \,|\, \hat{X}^{\hat{J}}(0) = X^J(t_1)\right] \,|\, X^J(0) = x\right]$$
$$= \mathbb{E}\left[g_1(J(t_1))\hat{\mathbb{E}}\left[(1 - X^J(t_1))^{\hat{R}_{T-t_1}^{\hat{J}}((T-t_1)-)} \,|\, \hat{R}_{T-t_1}^{\hat{J}}(0-) = n\right] \,|\, X^J(0) = x\right],$$

where the subordinator $\hat{J}$ is defined via $\hat{J}(h) := J(t_1 + h) - J(t_1)$. The processes $\hat{X}^{\hat{J}}$ and $\hat{R}_{T-t_1}^{\hat{J}}$ are independent copies of X^J and $R_{T-t_1}^J$, which are driven by $\hat{J}$ (which is in turn independent of $(J(u))_{u \in [0,t_1]}$). Using first Fubini's theorem, then Equation (2.7) in Theorem 2.3 and the fact

that 0 and t_1 are almost surely continuity times for J, we find that the last expression equals

$$\hat{\mathbb{E}}\left[\mathbb{E}\left[g_1(J(t_1))(1-X^J(t_1))^{\hat{R}^{\hat{J}}_{T-t_1}((T-t_1)-)}\mid X^J(0)=x\right]\mid\hat{R}^{\hat{J}}_{T-t_1}(0-)=n\right]$$

$$=\hat{\mathbb{E}}\left[\mathbb{E}\left[g_1(J(t_1))(1-x)^{R^J_{t_1}(t_1-)}\mid R^J_{t_1}(0-)=\hat{R}^{\hat{J}}_{T-t_1}((T-t_1)-)\right]\mid\hat{R}^{\hat{J}}_{T-t_1}(0-)=n\right].$$

The proof of the claim for $k=1$ is achieved using the Markov property for R^J_T in the (backward) interval $[0, T-t_1]$. Let us now assume that the claim is true up to $k-1$. We proceed as before to prove that the claim holds for k. For $j \in [k]$, define $G^\omega_{j,k}(z) := \prod_{i=j}^{k} g_i(z + \omega(t_i - t_1))$. Using the Markov property for X^J in $[0, t_1]$ followed by the inductive step, we obtain

$$\mathbb{E}\left[(1-X^J(T))^n \prod_{i=1}^{k} g_i(J(t_i))|X^J(0)=x\right]$$

$$=\mathbb{E}\left[g_1(J(t_1))\hat{\mathbb{E}}\left[(1-\hat{X}^{\hat{J}}(T-t_1))^n G^{\hat{J}}_{2,k}(J(t_1))|\hat{X}^{\hat{J}}(0)=X^J(t_1)\right]\mid X^J(0)=x\right]$$

$$=\mathbb{E}\left[g_1(J(t_1))\hat{\mathbb{E}}\left[(1-X^J(t_1))^{\hat{R}^{\hat{J}}_{T-t_1}((T-t_1)-)} G^{\hat{J}}_{2,k}(J(t_1))|\hat{R}^{\hat{J}}_{T-t_1}(0-)=n\right]|X^J(0)=x\right].$$

By Fubini's theorem, the reinforced duality Equation (2.7), and the fact that 0 and t_1 are almost surely continuity times for J, the last expression equals

$$\hat{\mathbb{E}}\left[\mathbb{E}\left[(1-X^J(t_1))^{\hat{R}^{\hat{J}}_{T-t_1}((T-t_1)-)}G^{\hat{J}}_{1,k}(J(t_1))|X^J(0)=x\right]|\hat{R}^{\hat{J}}_{T-t_1}(0-)=n\right]$$

$$=\hat{\mathbb{E}}\left[\mathbb{E}\left[(1-x)^{R^J_{t_1}(t_1-)}G^{\hat{J}}_{1,k}(J(t_1))|R^J_{t_1}(0-)=\hat{R}^{\hat{J}}_{T-t_1}((T-t_1)-)\right]|\hat{R}^{\hat{J}}_{T-t_1}(0-)=n\right].$$

The result follows from applying the Markov property for R^J_T in the (backward) interval $[0, T-t_1]$. $\qquad\square$

Proof of Theorem 2.4(2) *(asymptotic type frequency).* Let ω be such that Equation (2.9) holds between $-\tau$ and 0. In particular,

$$\mathbb{E}\left[(1-X^\omega(0))^n|X^\omega(-\tau)=x\right]=\mathbb{E}\left[(1-x)^{R^\omega_0(\tau-)}|R^\omega_0(0-)=n\right]. \tag{6.3}$$

Since we assume that $\theta > 0$ and $\nu_0, \nu_1 \in (0, 1)$, the right-hand side converges to $\Pi_n(\omega)$ (defined in (2.10)), which proves that the moment of order n of $1 - X^\omega(0)$ conditionally on $\{X^\omega(-\tau) = x\}$ converges to $\Pi_n(\omega)$. Since we are dealing with random variables supported on [0, 1], the convergence of the positive integer moments proves the convergence in distribution and the fact that the limit distribution $\mathcal{L}^\omega$ satisfies (2.13).

It remains to prove (2.14). For $\upsilon \in \mathcal{M}_1(\mathbb{N}_0^\dagger)$ with finite support, let υ_s^ω denote the distribution of $R^\omega_0(s-)$ given that $R^\omega_0(0-) \sim \upsilon$. Let $T^\omega_{0,\dagger}$ be the absorption time of R^ω_0 at $\{0, \dagger\}$. Note that $T^\omega_{0,\dagger}$ is stochastically bounded by an exponential random variable with parameter $\theta\nu_0$. Therefore,

$$\upsilon_\tau^\omega(\mathbb{N}) = \mathbb{P}_\upsilon\left(R^\omega_0(\tau-) \in \mathbb{N}\right) = \mathbb{P}_\upsilon\left(T^\omega_{0,\dagger} > \tau\right) \le e^{-\theta\nu_0\tau}.$$

Hence, we have

$$\mathbb{P}_\upsilon\left(\exists s \ge 0 \text{ s.t. } R^\omega_0(s) = 0\right) = \upsilon_\tau^\omega(\{0\}) + \sum_{k \ge 1} \mathbb{P}_\upsilon\left(R^\omega_0(\tau-) = k \text{ \& } \exists s \ge \tau \text{ s.t. } R^\omega_0(s) = 0\right)$$

$$\le \upsilon_\tau^\omega(\{0\}) + \upsilon_\tau^\omega(\mathbb{N}) \le \upsilon_\tau^\omega(\{0\}) + e^{-\theta\nu_0\tau}.$$

Thus, we obtain

$$\upsilon_\tau^\omega(\{0\}) \leq \mathbb{P}_\upsilon\big(\exists s \geq 0 \text{ s.t. } R_0^\omega(s) = 0\big) \leq \upsilon_\tau^\omega(\{0\}) + e^{-\theta v_0 \tau}. \tag{6.4}$$

Similarly, we have

$$\mathbb{E}_\upsilon\left[(1-x)^{R_0^\omega(\tau-)}\right] = \sum_{k \geq 0}(1-x)^k \upsilon_\tau^\omega(\{k\}) \leq \upsilon_\tau^\omega(\{0\}) + \upsilon_\tau^\omega(\mathbb{N}) \leq \upsilon_\tau^\omega(\{0\}) + e^{-\theta v_0 \tau}.$$

Hence,

$$\upsilon_\tau^\omega(\{0\}) \leq \mathbb{E}_\upsilon[(1-x)^{R_0^\omega(\tau-)}] \leq \upsilon_\tau^\omega(\{0\}) + e^{-\theta v_0 \tau}. \tag{6.5}$$

Recall from Section 2.5 that $\Pi_n(\omega) := \mathbb{P}(\exists s \geq 0 \text{ s.t. } R_0^\omega(s) = 0 \mid R_0^\omega(0-) = n)$. Choosing $\upsilon = \delta_n$ in (6.4) and in (6.5) and subtracting both inequalities, we get

$$\left|\mathbb{E}\left[(1-x)^{R_0^\omega(\tau-)} \mid R_0^\omega(0-) = n\right] - \Pi_n(\omega)\right| \leq e^{-\theta v_0 \tau}.$$

This inequality together with (2.9) (i.e. the quenched moment duality) yields the desired result. □

Proof of Proposition 2.1. Let $\omega \in \mathbb{D}^\star$ be such that (2.9) holds. Let $J_\omega := J \otimes_{\tau_\star} \omega$. Consider the process X^{J_ω} in $[-\tau, 0]$ with $\tau > \tau_\star$. Using the Markov property, we obtain

$$\mathbb{E}\big[(1-X^{J_\omega}(0))^n \mid X^{J_\omega}(-\tau) = x\big]$$
$$= \int_0^1 \mathbb{E}\big[(1-X^\omega(0))^n \mid X^\omega(-\tau_\star) = y\big] \, \mathbb{P}(X(-\tau_\star) \in \mathrm{d}y \mid X(-\tau) = x),$$

where X is the solution to (1.3) with subordinator J. Combining the previous identity with (2.9) for X^ω in $(-\tau_\star, 0)$, and using the translation invariance of X, we obtain

$$\mathbb{E}\big[(1-X^{J_\omega}(0))^n \mid X^{J_\omega}(-\tau) = x\big] =$$
$$\int_0^1 \mathbb{E}\big[(1-y)^{R_0^\omega(\tau_\star-)} \mid R_0^\omega(0-) = n\big] \, \mathbb{P}(X(\tau-\tau_\star) \in \mathrm{d}y \mid X(0) = x).$$

Hence, letting $\tau \to \infty$ and using Theorem 2.4(1), we get

$$\lim_{\tau \to \infty} \mathbb{E}\left[(1-X^{J_\omega}(0))^n \mid X^{J_\omega}(-\tau) = x\right]$$
$$= \int_0^1 \mathbb{E}\left[(1-y)^{R_0^\omega(\tau_\star-)} \mid R_0^\omega(0-) = n\right] \, \mathbb{P}(X(\infty) \in \mathrm{d}y),$$

and the result follows from Equation (2.11) in Theorem 2.4(1). □

6.2. Quenched results related to Section 2.6

This section is devoted to the proof of Theorem 2.5(2), which describes the asymptotic behavior of the ancestral type distribution.

Lemma 6.1. *For all $T \geq 0$, $x \in [0, 1]$, and $\omega \in \mathbb{D}^\star$, we have*

$$h_T^\omega(x) = 1 - \mathbb{E}[(1-x)^{L_T^\omega(T-)} \mid L_T^\omega(0-) = 1]. \tag{6.6}$$

Proof. The proof is analogous to the proof of Lemma 5.1.   □

Now we proceed to prove Theorem 2.5(2).

Proof of Theorem 2.5(2) *(ancestral type distribution).* Recall that $\theta v_0 > 0$ by assumption. For $\mu \in \mathcal{M}_1(\mathbb{N})$, we denote by $\mu_T^\omega(\beta)$ the distribution of $L_T^\omega(\beta-)$ given that $L_T^\omega(0-) \sim \mu$. Let $t > s > 0$. Note that we have $\mu_t^\omega(t) = (\mu_t^\omega(t-s))_s^\omega(s)$, so

$$d_{TV}(\mu_t^\omega(t), \mu_s^\omega(s)) = d_{TV}((\mu_t^\omega(t-s))_s^\omega(s), \mu_s^\omega(s)), \tag{6.7}$$

where $d_{TV}(\mu_1, \mu_2)$ stands for the total variation distance between μ_1 and μ_2.

Assume now that $L_T^\omega(0-) \sim \mu$. By construction, L_T^ω jumps from any state i to the state 1 with rate $q^0(i, 1) \geq \theta v_0 > 0$ (see (2.16)). Let $\hat{L}_T^\omega$ be a process with initial distribution μ, evolving as L_T^ω, but jumping from i to 1 at rate $q^0(i, 1) - \theta v_0 \geq 0$. We decompose the dynamic of L_T^ω as follows: (1) L_T^ω evolves as $\hat{L}_T^\omega$ on $[0, \xi]$, where ξ is an independent exponential random variable with parameter θv_0, (2) at time ξ, L_T^ω jumps to the state 1 regardless of its current position, and (3) conditionally on ξ, L_T^ω has the same law on $[\xi, \infty)$ as an independent copy of $L_{T-\xi}^\omega$ started with one line. This idea allows us to couple L_T^ω to a copy of it, $\tilde{L}_T^\omega$, with starting law $\tilde{\mu}$, so that the two processes are equal on $[\xi, \infty)$. Since $L_T^\omega(T-) \sim \mu_T^\omega(T)$ and $\tilde{L}_T^\omega(T-) \sim \tilde{\mu}_T^T(\omega)$, we have

$$d_{TV}(\mu_T^\omega(T), \tilde{\mu}_T^\omega(T)) \leq \mathbb{P}\left(\tilde{L}_T^\omega(T-) \neq L_T^\omega(T-)\right) \leq \mathbb{P}\left(\xi > T\right) = e^{-\theta v_0 T}. \tag{6.8}$$

This, together with (6.7), implies that for any $\mu \in \mathcal{M}_1(\mathbb{N})$ and any $t > s > 0$,

$$d_{TV}(\mu_t^\omega(t), \mu_s^\omega(s)) \leq e^{-\theta v_0 s}. \tag{6.9}$$

In particular, $(\mu_t^\omega(t))_{t>0}$ is Cauchy as $t \to \infty$ for the total variation distance. Therefore, $(\mu_t^\omega(t))_{t>0}$ has a limit $\mu^\omega \in \mathcal{M}_1(\mathbb{N})$. Moreover, (6.8) implies that μ^ω does not depend on μ, and the first part of Theorem 2.5(2) is proved. The identity (2.19) then follows by Lemma 6.1.

Setting $s = T$ and letting $t \to \infty$ in (6.9) yields $d_{TV}(\mu^\omega, \mu_T^\omega(T)) \leq e^{-\theta v_0 T}$. Since $h^\omega(x) = 1 - \mathbb{E}\left[(1-x)^{Z_\infty^\omega}\right]$ and $h_T^\omega(x) = 1 - \mathbb{E}\left[(1-x)^{Z_T^\omega}\right]$, where $Z_\infty^\omega \sim (\delta_1)^\omega$ and $Z_T^\omega \sim (\delta_1)_T^\omega(T)$, we get

$$|h_T^\omega(x) - h^\omega(x)| \leq d_{TV}((\delta_1)^\omega, (\delta_1)_T^\omega(T)) \leq e^{-\theta v_0 T},$$

completing the proof.   □

7. Further quenched results for simple environments

In this section we provide, for simple environments, extensions and refinements of the results obtained in Sections 2.5 and 2.6 in the quenched setting. Recall the quenched diffusion X^ω defined in Section 2.3.

7.1. Extensions of quenched results in Section 2.5

First we extend the main quenched results in Section 2.5, which hold for almost every environment, to any simple environment.

Theorem 7.1. (Quenched moment duality for simple environments.) *The quenched moment duality* (2.9) *holds for any simple environment.*

The proof of Theorem 7.1 has two main ingredients: a moment duality between the jumps of the environment, and a moment duality at the jumps. These results are covered by the next two lemmas.

Lemma 7.1. (Quenched moment duality between the jumps.) *Let $0 \leq s < t \leq T$, and assume that ω has no jumps in (s,t). For all $x \in [0, 1]$ and $n \in \mathbb{N}$, we have*

$$\mathbb{E}\left[(1 - X^\omega(t-))^n \mid X^\omega(s) = x\right] = \mathbb{E}\left[(1 - x)^{R_T^\omega((T-s)-)} \mid R_T^\omega(T - t) = n\right].$$

Proof. In (s, t), the processes X^ω and R_t^ω evolve as in the annealed case with $\mu = 0$. Therefore, the result follows from applying Theorem 2.3 with $\mu = 0$. □

Lemma 7.2. (Quenched moment duality at jumps.) *Assume that $\omega \in \mathbb{D}^\star$ is simple and has a jump at time $t < T$. Then, for all $x \in [0, 1]$ and $n \in \mathbb{N}$, we have*

$$\mathbb{E}\left[(1 - X^\omega(t))^n \mid X^\omega(t-) = x\right] = \mathbb{E}\left[(1 - x)^{R_T^\omega(T-t)} \mid R_T^\omega((T - t) -) = n\right].$$

Proof. On the one hand, since $X^\omega(t) = X^\omega(t-) + X^\omega(t-)(1 - X^\omega(t-))\Delta\omega(t)$ almost surely, we have

$$\mathbb{E}\left[(1 - X^\omega(t))^n \mid X^\omega(t-) = x\right] = [1 - x(1 + (1 - x)\Delta\omega(t))]^n = [(1 - x)(1 - x\Delta\omega(t))]^n. \tag{7.1}$$

On the other hand, conditionally on $\{R_T^\omega((T - t) -) = n\}$, we have $R_T^\omega(T - t) \sim n + Y$ where $Y \sim \text{Bin}(n, \Delta\omega(t))$. Therefore,

$$\mathbb{E}\left[(1 - x)^{R_T^\omega(T-t)} \mid R_T^\omega((T - t) -) = n\right] = \mathbb{E}\left[(1 - x)^{n+Y}\right]$$
$$= (1 - x)^n \left[1 - \Delta\omega(t) + \Delta\omega(t)(1 - x)\right]^n = [(1 - x)(1 - x\Delta\omega(t))]^n. \tag{7.2}$$

The combination of (7.1) and (7.2) yields the result. □

Proof of Theorem 7.1. Let ω be a simple environment. Let $(t_i)_{i=1}^m$ be the increasing sequence of jump times of ω in $[0, T]$. Without loss of generality we assume that 0 and T are both jump times of ω. In particular, $t_1 = 0$ and $t_m = T$. Let $(X^\omega(s))_{s \in [0,T]}$ and $(R_T^\omega(\beta))_{\beta \in [0,T]}$ be independent realizations of the Wright–Fisher process and the line-counting process of the k-ASG, respectively. For $s \in [0, T]$, denote by $\mu_s^\omega(x, \cdot)$ and $\bar{\mu}_s^\omega(x, \cdot)$ the laws of $X^\omega(s)$ and $X^\omega(s -)$, respectively, given that $X^\omega(0) = x$. Partitioning with respect to the values of $X^\omega(t_m -)$ and using Lemma 7.2 at $t = T$, we get

$$\mathbb{E}\left[(1 - X^\omega(T))^n \mid X^\omega(0) = x\right] = \int_0^1 \mathbb{E}\left[(1 - X^\omega(t_m))^n \mid X^\omega(t_m -) = y\right] \bar{\mu}_{t_m}^\omega(x, dy)$$

$$= \int_0^1 \mathbb{E}\left[(1 - y)^{R_T^\omega(0)} \mid R_T^\omega(0 -) = n\right] \bar{\mu}_{t_m}^\omega(x, dy)$$

$$= \mathbb{E}\left[(1 - X^\omega(t_m -))^{R_T^\omega(0)} \mid R_T^\omega(0 -) = n, X^\omega(0) = x\right] =: I_T^\omega(x, n).$$

For $t < T$, set $q_{n,k}^{\omega}(t) := \mathbb{P}\left(R_T^{\omega}(t) = k \mid R_T^{\omega}(0-) = n\right)$. Partitioning with respect to the values of $X^{\omega}(t_{m-1})$ and $R_T^{\omega}(0)$, and using Lemma 7.1, we get

$$
\begin{aligned}
I_T^{\omega}(x, n) &= \sum_{k \in \mathbb{N}_0^{\dagger}} \mathbb{E}\left[(1 - X^{\omega}(t_m-))^k \mid X^{\omega}(0) = x\right] q_{n,k}^{\omega}(0) \\
&= \sum_{k \in \mathbb{N}_0^{\dagger}} q_{n,k}^{\omega}(0) \int_0^1 \mathbb{E}\left[(1 - X^{\omega}(t_m-))^k \mid X^{\omega}(t_{m-1}) = y\right] \mu_{t_{m-1}}^{\omega}(x, \mathrm{d}y) \\
&= \sum_{k \in \mathbb{N}_0^{\dagger}} q_{n,k}^{\omega}(0) \int_0^1 \mathbb{E}\left[(1 - y)^{R_T^{\omega}((T-t_{m-1})-)} \mid R_T^{\omega}(0) = k\right] \mu_{t_{m-1}}^{\omega}(x, \mathrm{d}y) \\
&= \mathbb{E}\left[(1 - X^{\omega}(t_{m-1}))^{R_T^{\omega}((T-t_{m-1})-)} \mid R_T^{\omega}(0-) = n, X^{\omega}(0) = x\right].
\end{aligned}
$$

If $m = 2$, the proof of (2.9) is already complete. If $m > 2$, we continue as follows. Partitioning with respect to the values of $R_T^{\omega}((T - t_{m-1}) -)$ and of $X^{\omega}(t_{m-1} -)$, and using Lemma 7.2, we obtain

$$
\begin{aligned}
I_T^{\omega}(x, n) &= \sum_{k \in \mathbb{N}_0^{\dagger}} \mathbb{E}\left[(1 - X^{\omega}(t_{m-1}-))^k \mid X^{\omega}(0) = x\right] q_{n,k}^{\omega}((T - t_{m-1}) -) \\
&= \sum_{k \in \mathbb{N}_0^{\dagger}} q_{n,k}^{\omega}((T - t_{m-1}) -) \int_0^1 \mathbb{E}\left[(1 - X^{\omega}(t_{m-1}-))^k \mid X^{\omega}(t_{m-1} -) = y\right] \bar{\mu}_{t_{m-1}}^{\omega}(x, \mathrm{d}y) \\
&= \sum_{k \in \mathbb{N}_0^{\dagger}} q_{n,k}^{\omega}((T - t_{m-1}) -) \int_0^1 \mathbb{E}\left[(1 - y)^{R_T^{\omega}(T-t_{m-1})} \mid R_T^{\omega}((T - t_{m-1}) -) = k\right] \bar{\mu}_{t_{m-1}}^{\omega}(x, \mathrm{d}y) \\
&= \mathbb{E}\left[(1 - X^{\omega}(t_{m-1} -))^{R_T^{\omega}(T-t_{m-1})} \mid R_T^{\omega}(0-) = n, X^{\omega}(0) = x\right].
\end{aligned}
$$

Iterating this procedure, using successively Lemma 7.1 and Lemma 7.2 (the first one is applied on the intervals (t_{i-1}, t_i), while the second one is applied at the times t_i), we finally obtain

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

which ends the proof. $\square$

Theorem 7.2. (Quenched asymptotic type frequency for simple environments) *The statement of Theorem* 2.4(2) *holds for any simple environment.*

Proof. The proof is analogous to the proof of Theorem 2.4(2), but using Theorem 7.1 instead of Theorem 2.3. $\square$

Refinements for $\sigma = 0$. Under this additional assumption, we provide a more explicit expression for $\Pi_n(\omega)$ (defined in (2.10)). This is possible thanks to the following explicit diagonalization of $Q_{\dagger}^0$ (the transition matrix of R under the null environment).

Lemma 7.3. *Assume that* $\sigma = 0$, *and set* $\lambda_k^{\dagger} := -q_{\dagger}^0(k, k)$ *for* $k \in \mathbb{N}_0^{\dagger}$ *and* $\gamma_k^{\dagger} := q_{\dagger}^0(k, k-1)$ *for* $k \in \mathbb{N}$, *where* $q_{\dagger}^{\mu}(\cdot, \cdot)$ *is defined in* (2.6). *In addition, let* $D_{\dagger}$ *be the diagonal matrix with diagonal entries* $(-\lambda_i^{\dagger})_{i \in \mathbb{N}_0^{\dagger}}$, *and let* $U_{\dagger} := (u_{i,j}^{\dagger})_{i,j \in \mathbb{N}_0^{\dagger}}$ *and* $V_{\dagger} := (v_{i,j}^{\dagger})_{i,j \in \mathbb{N}_0^{\dagger}}$ *be defined via*

$$u_{i,j}^\dagger := \begin{cases} \prod\limits_{\ell=j+1}^{i} \left(\dfrac{\gamma_\ell^\dagger}{\lambda_\ell^\dagger - \lambda_j^\dagger} \right) & \textit{for } i \in \mathbb{N}_0 \ \& \ j \in [i]_0, \\[4pt] 0 & \textit{for } i \in \mathbb{N}_0 \ \& \ j > i, \text{ or } i = \dagger \ j \in \mathbb{N}_0, \\[4pt] \theta v_0 \sum\limits_{k=1}^{i} \dfrac{k}{\lambda_k^\dagger} \prod\limits_{\ell=k+1}^{i} \dfrac{\gamma_\ell^\dagger}{\lambda_\ell^\dagger} & \textit{for } i \in \mathbb{N}_0 \ \& \ j = \dagger, \\[4pt] 1 & \textit{for } i = j = \dagger, \end{cases} \tag{7.3}$$

$$v_{i,j}^\dagger := \begin{cases} \prod\limits_{\ell=j}^{i-1} \left(\dfrac{-\gamma_{\ell+1}^\dagger}{\lambda_i^\dagger - \lambda_\ell^\dagger} \right) & \textit{for } i \in \mathbb{N}_0 \ \& \ j \in [i]_0, \\[4pt] 0 & \textit{for } i \in \mathbb{N}_0 \ \& \ j > i, \text{ or } i = \dagger \ \ j \in \mathbb{N}_0, \\[4pt] \dfrac{-\theta v_0}{\lambda_i^\dagger} \sum\limits_{k=1}^{i} k \prod\limits_{\ell=k}^{i-1} \left(\dfrac{-\gamma_{\ell+1}^\dagger}{\lambda_i^\dagger - \lambda_\ell^\dagger} \right) & \textit{for } i \in \mathbb{N}_0 \ \& \ j = \dagger, \\[4pt] 1 & \textit{for } i = j = \dagger, \end{cases} \tag{7.4}$$

with the convention that an empty sum equals 1 *and an empty product equals* 1. *Then we have* $Q_\dagger^0 = U_\dagger D_\dagger V_\dagger$ *and* $U_\dagger V_\dagger = V_\dagger U_\dagger = I$, *where* I *denotes the identity matrix.*

Proof of Lemma 7.3. For any $i \in \mathbb{N}_0^\dagger$, let $e_i := (e_{i,j})_{j \in \mathbb{N}_0^\dagger}$ be the vector defined via $e_{i,i} := 1$ and $e_{i,j} := 0$ for $j \neq i$. Order $\mathbb{N}_0^\dagger$ as $\{\dagger, 0, 1, 2, \ldots\}$, so that the matrix $(Q_\dagger^0)^\top$ is upper triangular with diagonal elements $(-\lambda_\dagger, -\lambda_0, -\lambda_1, -\lambda_2, \ldots)$. For $n \in \mathbb{N}_0^\dagger$, let $v_n \in \mathrm{Span}\{e_i : i \in [n]_0 \cup \{\dagger\}\}$ be the eigenvector of $(Q_\dagger^0)^\top$ associated with the eigenvalue $-\lambda_n$, normalized so that its coordinate with respect to e_n is 1. It is not difficult to see that these eigenvectors exist and that we have $v_\dagger = e_\dagger$ and $v_0 = e_0$. For $n \geq 1$, writing $v_n = c_\dagger e_\dagger + c_0 e_0 + \ldots + c_{n-1} e_{n-1} + e_n$ and multiplying by $\frac{1}{-\lambda_n}(Q_\dagger^0)^\top$ on both sides, we obtain another expression for v_n as a linear combination of $e_\dagger, e_0, \ldots, e_{n-1}, e_n$. Identifying the two expressions, we obtain that $c_k = v_{n,k}^\dagger$, for $k \leq n-1$. In particular, we have

$$v_n = v_{n,\dagger}^\dagger e_\dagger + v_{n,0}^\dagger e_0 + \cdots + v_{n,n-1}^\dagger e_{n-1} + v_{n,n}^\dagger e_n.$$

Proceeding in a similar way, one obtains that

$$e_n = u_{n,\dagger}^\dagger v_\dagger + u_{n,0}^\dagger v_0 + \cdots + u_{n,n-1}^\dagger v_{n-1} + u_{n,n}^\dagger v_n.$$

We thus get that $V_\dagger^\top U_\dagger^\top = U_\dagger^\top V_\dagger^\top = I$ and $(Q_\dagger^0)^\top = V_\dagger^\top D_\dagger U_\dagger^\top$ (the matrix products are well-defined, because they involve sums of finitely many non-zero terms). This ends the proof. $\qquad\square$

Now, consider the polynomials $S_k^\dagger$, $k \in \mathbb{N}_0$, defined via

$$S_k^\dagger(x) := \sum_{i=0}^{k} v_{k,i}^\dagger x^i, \quad x \in [0, 1]. \tag{7.5}$$

For $z \in (0, 1)$, define the matrices $\mathcal{B}(z) := (\mathcal{B}_{i,j}(z))_{i,j \in \mathbb{N}_0^\dagger}$ and $\Phi^\dagger(z) := (\Phi_{i,j}^\dagger(z))_{i,j \in \mathbb{N}_0^\dagger}$ via

$$\mathcal{B}_{i,j}(z) := \begin{cases} \mathbb{P}(i + B_i(z) = j) & \text{for } i, j \in \mathbb{N}, \\ 1 & \text{for } i = j \in \{0, \dagger\}, \qquad \& \quad \Phi^\dagger(z) := U_\dagger^\top \mathcal{B}(z)^\top V_\dagger^\top, \\ 0 & \text{otherwise,} \end{cases} \tag{7.6}$$

where $B_i(z) \sim \text{Bin}(i, z)$. We will see in the proof of Theorem 7.3 that $\Phi^\dagger(z)$ is well-defined.

Theorem 7.3. *Assume that $\sigma = 0$, $\theta > 0$, and $\nu_0, \nu_1 \in (0, 1)$. Let ω be a simple environment. Denote by $N := N(\tau)$ the number of jumps of ω in $(-\tau, 0)$, and let $(T_i)_{i=1}^N$ be the sequence of the jump times in decreasing order; set $T_0 := 0$. For any $m \in [N]$, define the matrix $A_m^\dagger(\omega) := (A_{i,j}^{\dagger,m}(\omega))_{i,j \in \mathbb{N}_0^\dagger}$ via*

$$A_m^\dagger(\omega) := \Phi^\dagger(\Delta\omega(T_m)) \exp\left((T_{m-1} - T_m)D_\dagger\right). \tag{7.7}$$

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

$$\mathbb{E}\left[(1 - X^\omega(0))^n \mid X^\omega(-\tau) = x\right] = \sum_{k=0}^{n2^N} C_{n,k}^\dagger(\omega, \tau) S_k^\dagger(1-x), \tag{7.8}$$

where the matrix $C^\dagger(\omega, \tau) := (C_{n,k}^\dagger(\omega, \tau))_{k,n \in \mathbb{N}_0^\dagger}$ is given by

$$C^\dagger(\omega, \tau) := U_\dagger \left[A_N^\dagger(\omega)A_{N-1}^\dagger(\omega) \cdots A_1^\dagger(\omega)\right]^\top \exp\left((T_N + \tau)D_\dagger\right), \tag{7.9}$$

with the convention that an empty product of matrices is the identity matrix. Moreover, for all $n \in \mathbb{N}$,

$$\Pi_n(\omega) = C_{n,0}^\dagger(\omega, \infty) := \lim_{\tau \to \infty} C_{n,0}^\dagger(\omega, \tau) = \lim_{\tau \to \infty} \left(U_\dagger \left[A_{N(\tau)}^\dagger(\omega) \cdots A_1^\dagger(\omega)\right]^\top\right)_{n,0}, \tag{7.10}$$

where the previous limits are well-defined.

Proof. Let us first show that the matrix products in (7.6), (7.7), and (7.9) are well-defined and that $C_{n,k}^\dagger(\omega, \tau) = 0$ for all $k > n2^N$. To this end, order $\mathbb{N}_0^\dagger$ as $\{\dagger, 0, 1, 2, \ldots\}$, so that the matrices $U_\dagger^\top$ and $V_\dagger^\top$ are upper triangular. Note also that $\mathcal{B}_{j,i}(z) = 0$ for $i > 2j$. Therefore, for any $n \in \mathbb{N}$ and any $v = (v_i)_{i \in \mathbb{N}_0^\dagger}$ such that $v_i = 0$ for all $i > n$, the vector $\tilde{v} := U_\dagger^\top(\mathcal{B}(z)^\top(V_\dagger^\top v))$ is well-defined and satisfies $\tilde{v}_i = 0$ for all $i > 2n$. It follows that the matrix $\Phi^\dagger(z)$ in (7.6) is well-defined. Moreover, since $\exp\left((T_{m-1} - T_m)D_\dagger\right)$ is diagonal, the product defining the matrix $A_m^\dagger(\omega)$ in (7.7) is also well-defined. Furthermore, for any $n \in \mathbb{N}$ and any vector $v = (v_i)_{i \in \mathbb{N}_0^\dagger}$ such that $v_i = 0$ for all $i > n$, the vector $\tilde{v} := A_m^\dagger(\omega)v$ satisfies $\tilde{v}_i = 0$ for all $i > 2n$. In particular, for any $m \geq 1$, the product $\exp\left(-(T_N + \tau)D_\dagger\right)A_m^\dagger(\omega)A_{m-1}^\dagger(\omega) \cdots A_1^\dagger(\omega)U_\dagger^\top$ is well-defined. Additionally, for $n \geq 1$ and a vector $v = (v_i)_{i \in \mathbb{N}_0^\dagger}$ such that $v_i = 0$ for all $i > n$, the vector $\tilde{v} := \exp\left(-(T_N + \tau)D_\dagger\right)A_m^\dagger(\omega)A_{m-1}^\dagger(\omega) \cdots A_1^\dagger(\omega)U_\dagger^\top v$ satisfies $\tilde{v}_i = 0$ for all $i > 2^m n$. Transposing, we see that the matrix $C^\dagger(\omega, \tau)$ in (7.9) is well-defined and satisfies $C_{n,k}^\dagger(\omega, \tau) = 0$ for all $k > n2^N$.

For $s > 0$, define the stochastic matrix $\mathcal{P}_s^\dagger(\omega) := (p_{i,j}^\dagger(\omega, s))_{i,j \in \mathbb{N}_0^\dagger}$ via

$$p_{i,j}^\dagger(\omega, s) := \mathbb{P}(R_0^\omega(s-) = j \mid R_0^\omega(0-) = i).$$

Hence, defining $\rho(y) := (y^i)_{i \in \mathbb{N}_0^\dagger}$, $y \in [0, 1]$ (with the convention $y^\dagger := 0$), we obtain

$$\mathbb{E}[y^{R_0^\omega(\tau-)} \mid R_0^\omega(0-) = n] = (\mathcal{P}_\tau^\dagger(\omega)\rho(y))_n = (\mathcal{P}_\tau^\dagger(\omega)U_\dagger S_\dagger(y))_n, \tag{7.11}$$

where we used that $\rho(y) = U_\dagger S_\dagger(y)$ with $S_\dagger(y) := (S_k^\dagger(y))_{k \in \mathbb{N}_0^\dagger}$. Thus, Theorem 7.1 and Equation (7.11) yield

$$\mathbb{E}[(1 - X^\omega(0))^n \mid X^\omega(-\tau) = x] = \sum_{k=0}^\infty \left(\mathcal{P}_\tau^\dagger(\omega) U_\dagger \right)_{n,k} S_k^\dagger(1 - x). \tag{7.12}$$

Now, consider the semigroup $M_\dagger := (M_\dagger(s))_{s \geq 0}$ of the line-counting process of the k-ASG in the null environment, which is defined via $M_\dagger(s) := \exp(s Q_\dagger^0)$. Thanks to Lemma 7.3, $M_\dagger(\beta) = U_\dagger E_\dagger(\beta) V_\dagger$, where $E_\dagger(\beta)$ is the diagonal matrix with diagonal entries $(e^{-\lambda_j^\dagger \beta})_{j \in \mathbb{N}_0^\dagger}$.

Assume first that $N(\tau) = 0$ (i.e. ω has no jumps in $[-\tau, 0]$). In this case, we have

$$\mathcal{P}_\tau^\dagger(\omega) U_\dagger = M_\dagger(\tau) U_\dagger = U_\dagger E_\dagger(\tau) V_\dagger U_\dagger = U_\dagger E_\dagger(\tau) = C^\dagger(\omega, \tau),$$

where we used that $V_\dagger U_\dagger = I$. Hence, (7.8) follows from (7.12).

Assume now that $N(\tau) \geq 1$ (i.e. ω has at least one jump in $[-\tau, 0]$). Disintegrating with respect to the values of $R_0^\omega((-T_i)-)$ and $R_0^\omega(-T_i)$, $i \in [N]$, we get

$$\mathcal{P}_\tau^\dagger(\omega) = M_\dagger(-T_1) \mathcal{B}(\Delta\omega(T_1)) M_\dagger(T_1 - T_2) \mathcal{B}(\Delta\omega(T_2)) \cdots \mathcal{B}(\Delta\omega(T_N)) M_\dagger(T_N + \tau). \tag{7.13}$$

Using this, the relation $M_\dagger(\beta) = U_\dagger E_\dagger(\beta) V_\dagger$, the definition of the matrices $\Phi^\dagger$ and $A_i^\dagger$ (see (7.6) and (7.7)), and the fact that $V_\dagger U_\dagger = I$, we obtain

$$
\begin{aligned}
&\mathcal{P}_\tau^\dagger(\omega) U_\dagger \\
&= U_\dagger E_\dagger(-T_1) \Phi^\dagger(\Delta\omega(T_1))^\top E_\dagger(T_1 - T_2) \Phi^\dagger(\Delta\omega(T_2))^\top \cdots \Phi^\dagger(\Delta\omega(T_N))^\top E_\dagger(T_N + \tau) \\
&= U_\dagger A_1^\dagger(\omega)^\top A_2^\dagger(\omega)^\top \cdots A_N^\dagger(\omega)^\top E_\dagger(T_N + \tau) \\
&= U_\dagger \left[A_N^\dagger(\omega) A_{N-1}^\dagger(\omega) \cdots A_1^\dagger(\omega) \right]^\top E_\dagger(T_N + \tau) = C^\dagger(\omega, \tau), \tag{7.14}
\end{aligned}
$$

which proves (7.8) also in this case.

It remains to prove that $C_{n,0}^\dagger(\omega, \tau)$ converges to $\Pi_n(\omega)$ as $\tau \to \infty$. For $\omega = \mathbf{0}$ (i.e. the null environment), on the one hand (7.9) yields $C_{n,0}^\dagger(\omega, \tau) = e^{-\lambda_0^\dagger \tau} u_{n,0}^\dagger$, and on the other hand (7.11) together with $M_\dagger(\beta) = U_\dagger E_\dagger(\beta) V_\dagger$ and $V_\dagger U_\dagger = I$ yields

$$\mathbb{E}[y^{R_0^\mathbf{0}(\tau-)} \mid R_0^\mathbf{0}(0-) = n] = \sum_{k=0}^n e^{-\lambda_k^\dagger \tau} u_{n,k}^\dagger S_k^\dagger(y).$$

Since $\lambda_k^\dagger > 0$ for $k \in \mathbb{N}$ and $\lambda_0^\dagger = 0$, the desired convergence follows by letting $\tau \to \infty$ in the previous identity. For later use, note that we have $\Pi_n(\mathbf{0}) = u_{n,0}^\dagger$. The general case is a direct consequence of the following proposition. $\qquad\square$

Proposition 7.1. *Assume that $\sigma = 0$, $\theta > 0$, $v_0, v_1 \in (0, 1)$, and ω is a simple environment. We have*

$$\left| C_{n,0}^\dagger(\omega, \tau) - \Pi_n(\omega) \right| \leq e^{-\theta v_0 \tau}.$$

Proof. Let ω_τ be the environment that coincides with ω in $(-\tau, \infty)$ and that is constant and equal to $\omega(-\tau)$ in $(-\infty, -\tau]$ (which means that ω_τ has no jumps in $(-\infty, -\tau]$). Since

$\mathcal{P}_\tau^\dagger(\omega_\tau) = \mathcal{P}_\tau^\dagger(\omega)$ and ω_τ has no jumps in $(-\infty, -\tau]$, we obtain

$$\Pi_n(\omega_\tau) = \sum_{k \geq 0} p_{n,k}^\dagger(\omega_\tau, \tau) \, \mathbb{P}(\exists \beta \geq \tau \text{ s.t. } R_0^{\omega_\tau}(\beta) = 0 \mid R_0^{\omega_\tau}(\tau-) = k)$$

$$= \sum_{k \geq 0} p_{n,k}^\dagger(\omega, \tau) \, \Pi_k(\mathbf{0}) = \sum_{k \geq 0} p_{n,k}^\dagger(\omega, \tau) \, u_{k,0}^\dagger = (\mathcal{P}_\tau^\dagger(\omega) U_\dagger)_{n,0} = C_{n,0}^\dagger(\omega, \tau), \qquad (7.15)$$

where in the last line we used $\Pi_k(\mathbf{0}) = u_{k,0}^\dagger$ (from the end of the previous proof) and (7.14). Now combining (7.15) with (6.4) applied to ω_τ with $\upsilon = \delta_n$ yields

$$p_{n,0}^\dagger(\omega, \tau) = p_{n,0}^\dagger(\omega_\tau, \tau) \leq C_{n,0}^\dagger(\omega, \tau) \leq p_{n,0}^\dagger(\omega_\tau, \tau) + e^{-\theta \upsilon_0 \tau} = p_{n,0}^\dagger(\omega, \tau) + e^{-\theta \upsilon_0 \tau}. \quad (7.16)$$

Then, combining (6.4) applied to ω with $\upsilon = \delta_n$ and (7.16), we get

$$C_{n,0}^\dagger(\omega, \tau) - e^{-\theta \upsilon_0 \tau} \leq \Pi_n(\omega) \leq C_{n,0}^\dagger(\omega, \tau) + e^{-\theta \upsilon_0 \tau},$$

and the result follows. $\qquad \square$

Remark 7.1. If ω has no jumps in $(-\tau, 0)$, then $C^\dagger(\omega, \tau) = U_\dagger \exp(\tau D_\dagger)$. In particular, $\Pi_n(\mathbf{0}) = u_{n,0}^\dagger$.

Remark 7.2. Under the assumptions of Theorem 7.3, the Simpson index (see Remark 2.5) is given by

$$\mathbb{E}[\text{Sim}^\omega(\infty)] = \mathbb{E}[X^\omega(\infty)^2 + (1 - X^\omega(\infty))^2] = 1 - 2C_{1,0}^\dagger(\omega, \infty) + 2C_{2,0}^\dagger(\omega, \infty).$$

Remark 7.3. If ω is a simple periodic environment with period $T_p > 0$, then (7.10) can be written as $\Pi_n(\omega) = \lim_{m \to \infty} (U_\dagger B(\omega)^m)_{n,0}$ where $B(\omega) := [A_{N(T_p)}^\dagger(\omega) \cdots A_1^\dagger(\omega)]^\top$. As an application of Theorem 7.3 we obtain the following refinement of Proposition 2.1 for mixed environments composed of a pure-jump subordinator J and a simple environment ω (see Figure 6).

Proposition 7.2. *Assume that* $\sigma = 0$, $\theta > 0$, *and* $\upsilon_0, \upsilon_1 \in (0, 1)$. *For any* $\tau_\star > 0$, $n \in \mathbb{N}$, $x \in [0, 1]$, *and any simple environment* ω, *we have*

$$\lim_{\tau \to \infty} \mathbb{E}[(1 - X^{J \otimes \tau_\star \omega}(0))^n | X^{J \otimes \tau_\star \omega}(-\tau) = x] = \sum_{j=0}^{n2^N} \left(\sum_{k=j}^{n2^N} C_{n,k}^\dagger(\omega, \tau_\star) v_{k,j}^\dagger \right) \pi_j, \qquad (7.17)$$

where N denotes the number of jumps of ω *in* $[-\tau_\star, 0]$.

Proof. Let ω be a simple environment. Proceeding as in the proof of Proposition 2.1, but using Theorem 7.1 instead of Theorem 2.3, we obtain

$$\lim_{\tau \to \infty} \mathbb{E}\left[(1 - X^{J \otimes \tau_\star \omega}(0))^n | X^{J \otimes \tau_\star \omega}(-\tau) = x\right] = \mathbb{E}\left[\pi_{R_0^\omega(\tau_\star-)} \mid R_0^\omega(0-) = n\right].$$

Since $U^\dagger V^\dagger = I$ and the stochastic matrix $\mathcal{P}_{\tau_\star}^\dagger(\omega) := (p_{i,j}^\dagger(\omega, \tau_\star))_{i,j \in \mathbb{N}_0^\dagger}$ defined via

$$p_{i,j}^\dagger(\omega, \tau_\star) := \mathbb{P}(R_0^\omega(\tau_\star-) = j \mid R_0^\omega(0-) = i)$$

satisfies $C^\dagger(\omega, \tau_\star) = \mathcal{P}_{\tau_\star}^\dagger(\omega) U_\dagger$ (see (7.14)), the result follows. $\qquad \square$

7.2. Extensions of quenched results in Section 2.6

In this section we assume that $\sigma = 0$ and extend some of the quenched results stated in Section 2.6 concerning the ancestral type distribution for simple environments. The next result allows us to get rid of the condition $\theta \nu_0 > 0$ in Theorem 2.5(2).

Theorem 7.4. (Ancestral type distribution for simple environments.) *Assume that $\sigma = 0$, and let $\omega \in \mathbb{D}^\star$ be a simple environment with infinitely many jumps in $[0, \infty)$ and such that the distance between the successive jumps does not converge to* 0. *Then the statement of Theorem 2.5(2), except for Equation (2.20), remains true.*

Proof. The case $\theta \nu_0 > 0$ is already covered by Theorem 2.5(2). Assume now that $\theta \nu_0 = 0$, $\sigma = 0$, and that ω is as in the statement. For $\mu \in \mathcal{M}_1(\mathbb{N})$, we denote by $\mu_T^\omega(\beta)$ the distribution of $L_T^\omega(\beta -)$ given that $L_T^\omega(0 -) \sim \mu$. We claim that, for all $\mu, \tilde{\mu} \in \mathcal{M}_1(\mathbb{N})$,

$$d_{TV}(\mu_T^\omega(T), \tilde{\mu}_T^\omega(T)) \xrightarrow[T \to \infty]{} 0, \qquad \text{(Claim 5)}$$

where $d_{TV}(\mu_1, \mu_2)$ stands for the total variation distance between μ_1 and μ_2. If Claim 5 is true, the rest of the proof follows as in the proof of Theorem 2.5(2). In what follows we prove Claim 5.

Let $0 < T_1 < T_2 < \cdots$ be the sequence of the jump times of ω, and set $T_0 := 0$ for convenience. On (T_i, T_{i+1}), L_T^ω has transition rates given by $(q^0(k, j))_{k,j \in \mathbb{N}}$ (see (2.16)). 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 $(S_l)_{l \geq 2}$ be a sequence of independent exponential random variables with parameter $(l - 1) \theta \nu_1 + l(l - 1)/2$ and note that $S_l \sim H(l, l - 1)$ for $l \geq 2$. Using the Markov property, one can easily see that $H(k, 1)$ is equal in distribution to $\sum_{l=2}^{k} S_l$. Therefore, for any i such that $T_{i+1} < T$ and any $k \geq 1$, we have

$$\mathbb{P}\left(L_T^\omega((T - T_i) -) = 1 \mid L_T^\omega(T - T_{i+1}) = k\right) = \mathbb{P}\left(\sum_{l=2}^{k} S_l \leq T_{i+1} - T_i\right)$$

$$\geq \mathbb{P}\left(\sum_{l=2}^{\infty} S_l \leq T_{i+1} - T_i\right). \qquad (7.18)$$

Clearly $\sum_{l=2}^{\infty} \mathbb{E}[S_l] < \infty$; thus $S^\infty := \sum_{l=2}^{\infty} S_l < \infty$ almost surely. Moreover, since for $l \geq 2$ the support of S_l contains 0, the support of S^∞ contains 0 as well. In particular $q(s) := \mathbb{P}(\sum_{l=2}^{\infty} S_l \leq s)$ is positive for all $s > 0$. Thus, we get from (7.18) that

$$\mathbb{P}\left(L_T^\omega((T - T_i) -) = 1 \mid L_T^\omega(T - T_{i+1}) = k\right) \geq q(T_{i+1} - T_i), \quad k \geq 1. \qquad (7.19)$$

Let L_T^ω, $\tilde{L}_T^\omega$, and $L_{T_i}^\omega$, $i \geq 0$, be independent copies of the line-counting process of the pLD-ASG with environment ω (the subscript indicates the sampling time) and $L_T^\omega(0 -) \sim \mu$, $\tilde{L}_T^\omega(0 -) \sim \tilde{\mu}$, and $L_{T_i}^\omega(0 -) = 1$. Let

$$i(T) := \max\{i \in \mathbb{N}_0 : T_i < T, \ L_T^\omega((T - T_i) -) = \tilde{L}_T^\omega((T - T_i) -) = 1\},$$

with the convention that the maximum of an empty set is $-\infty$. Set $T_{-\infty} := -\infty$ for convenience. We define $(U_T^\omega(\beta))_{\beta \geq 0}$ and $(\tilde{U}_T^\omega(\beta))_{\beta \geq 0}$ by setting $U_T^\omega(\beta) := L_T^\omega(\beta)$ and $\tilde{U}_T^\omega(\beta) := \tilde{L}_T^\omega(\beta)$ for $\beta < T - T_{i(T)}$ and $U_T^\omega(\beta) := \tilde{U}_T^\omega(\beta) = L_{T_{i(T)}}^\omega(\beta - (T - T_{i(T)}))$ for

$\beta \geq T - T_{i(T)}$. Note that U_T^ω and $\tilde{U}_T^\omega$ have the same distributions as L_T^ω and $\tilde{L}_T^\omega$, respectively. In particular, we have $U_T^\omega(T-) \sim \mu_T^\omega(T)$ and $\tilde{U}_T^\omega(T-) \sim \tilde{\mu}_T^\omega(T)$. Moreover, we have $U_T(\omega, \beta) = \tilde{U}_T(\omega, \beta)$ for all $\beta \geq T - T_{i(T)}$. Therefore,

$$d_{TV}(\mu_T^\omega(T), \tilde{\mu}_T^\omega(T)) \leq \mathbb{P}\left(U_T^\omega(T-) \neq \tilde{U}_T^\omega(T-)\right) \leq \mathbb{P}\left(i(T) = -\infty\right). \tag{7.20}$$

Let $N(T)$ be the number of jumps of ω in $[0, T]$. According to (7.19), for $k_1, k_2 \geq 1$ with $k_1 \neq k_2$ we have

$$\mathbb{P}\left(L_T^\omega((T - T_i)-) = 1, \ \tilde{L}_T^\omega((T - T_i)-) = 1 \mid L_T^\omega(T - T_{i+1}) = k_1, \tilde{L}_T^\omega(T - T_{i+1}) = k_2\right)$$
$$\geq q(T_{i+1} - T_i)^2.$$

Therefore, using (7.20), we obtain

$$d_{TV}(\mu_T^\omega(T), \tilde{\mu}_T^\omega(T)) \leq \mathbb{P}\left(I_0(T) = -\infty\right) \leq \prod_{i=1}^{N(T)} \left(1 - q(T_i - T_{i-1})^2\right) =: \varphi_\omega(T).$$

Note that φ_ω does not depend on μ and $\tilde{\mu}$. Recall that by assumption the sequence of jump 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 $T_{i+1} - T_i > \epsilon$. Thus, the number of factors smaller than $1 - q(\epsilon)^2 < 1$ in the product defining $\varphi_\omega(T)$ converges to infinity as $T \to \infty$. We deduce that $\varphi_\omega(T) \to 0$ as $T \to \infty$, which proves Claim 5, concluding the proof.   □

The following diagonalization of Q^0 (the transition matrix of the process L under the null environment) will allow us to obtain a more explicit expression for $h_T^\omega(x)$.

Lemma 7.4. *Assume that $\sigma = 0$, and for $k \in \mathbb{N}$ set $\lambda_k := -q^0(k, k)$ and $\gamma_k := q^0(k, k-1)$, where $q^\mu(\cdot, \cdot)$ is defined in (2.16). In addition, make the following definitions:*

(i) *Let D be the diagonal matrix with diagonal entries $(-\lambda_i)_{i \in \mathbb{N}}$.*

(ii) *Let $U := (u_{i,j})_{i,j \in \mathbb{N}}$, where, for all $i \in \mathbb{N}$, $u_{i,i} := 1$; $u_{i,j} := 0$ for $j > i$; 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*

$$u_{i,j} := \frac{1}{\lambda_i - \lambda_j}\left(\gamma_i u_{i-1,j} + \theta v_0\left(\sum_{l=j}^{i-2} u_{l,j}\right)\right). \tag{7.21}$$

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

$$v_{i,j} := \frac{-1}{(\lambda_i - \lambda_j)}\left[\left(\sum_{l=j+2}^{i} v_{i,l}\right)\theta v_0 + v_{i,j+1}\gamma_{j+1}\right]. \tag{7.22}$$

(We adopt the convention that an empty sum equals 0.) Then we have $Q = UDV$ and $UV = VU = I$, where I denotes the identity matrix.

Proof. The proof is analogous to the proof of Lemma 7.3. □

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

$$S_k(x) := \sum_{i=1}^{k} v_{k,i} x^i. \tag{7.23}$$

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

$$\mathcal{B}_{i,j}(z) := \mathbb{P}(i + B_i(z) = j), \quad i, j \in \mathbb{N}, \quad \text{and} \quad \Phi(z) := U^\top \mathcal{B}(z)^\top V^\top, \tag{7.24}$$

where $B_i(z) \sim \text{Bin}(i, z)$. The fact that the matrix product defining $\Phi(z)$ is well-defined can be justified similarly as in the proof of Theorem 7.3. The same is true for the matrix products in (7.25) and (7.27).

Theorem 7.5. *Assume that $\sigma = 0$, and let ω 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 ω in $(0,T)$, and let $(T_i)_{i=1}^{N}$ be the sequence of the jump times 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*

$$A_m(\omega) := \exp\left((T_m - T_{m-1})D\right) \Phi(\Delta\omega(T_m)). \tag{7.25}$$

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

$$h_T^\omega(x) = 1 - \sum_{k=1}^{2^N} C_{1,k}(\omega, T) S_k(1 - x), \tag{7.26}$$

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

$$C(\omega, T) := U \exp((T - T_N)D) \left[A_1(\omega)A_2(\omega)\cdots A_N(\omega)\right]^\top. \tag{7.27}$$

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

$$h^\omega(x) = 1 - \sum_{k=1}^{\infty} C_{1,k}(\omega, \infty) S_k(1 - x), \tag{7.28}$$

where the series in (7.28) is convergent and where

$$C_{1,k}(\omega, \infty) := \lim_{m\to\infty} \left(U \left[A_1(\omega)A_2(\omega)\cdots A_m(\omega)\right]^\top\right)_{1,k}, \tag{7.29}$$

and the above limit is well-defined.

Proof. 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 $M(s) := \exp(sQ^0)$; i.e. M is the semigroup of L^0. Let $T_1 < T_2 < \cdots < T_N$ be the sequence of jump times of ω in $[0, T]$. Disintegrating with respect to the values of $L_T^\omega((T - T_i)-)$ and $L_T^\omega(T - T_i)$, $i \in [N]$, we obtain

$$\mathcal{P}_T^T(\omega) = M(T - T_N)\mathcal{B}(\Delta\omega(T_N))M(T_N - T_{N-1})\mathcal{B}(\Delta\omega(T_{N-1}))\cdots\mathcal{B}(\Delta\omega(T_1))M(T_1). \quad (7.30)$$

In addition,

$$\mathbb{E}[y^{L_T^\omega(T-)} \mid L_T^\omega(0-) = n] = (\mathcal{P}_T^T(\omega)\rho(y))_n, \quad \text{where} \quad \rho(y) := (y^i)_{i \in \mathbb{N}}. \quad (7.31)$$

Thanks to Lemma 7.4, we have $M(\beta) = UE(\beta)V$, where $E(\beta)$ is the diagonal matrix with diagonal entries $(e^{-\lambda_j \beta})_{j \in \mathbb{N}}$. Moreover, $\rho(y) = US(y)$, where $S(y) := (S_k(y))_{k \in \mathbb{N}}$. Using this together with Equation (7.30) and the relations $M(\beta) = UE(\beta)V$ and $VU = I$, we obtain

$$\mathcal{P}_T^T(\omega)\rho(y)$$
$$= UE(T - T_N)\Phi(\Delta\omega(T_N))^\top E(T_N - T_{N-1})\Phi(\Delta\omega(T_{N-1}))^\top \cdots \Phi(\Delta\omega(T_1))^\top E(T_1)S(y). \quad (7.32)$$

Thus, using the definition of the matrices $A_i(\omega)$, we get

$$\mathcal{P}_T^T(\omega)\rho(y) = UE(T - T_N)A_N(\omega)^\top A_{N-1}(\omega)^\top \cdots A_1(\omega)^\top S(y)$$
$$= UE(T - T_N)[A_1(\omega)A_2(\omega)\cdots A_N(\omega)]^\top S(y) = C(\omega, T)S(y).$$

Now, using the previous identity, Lemma 6.1, and Equation (7.31), we obtain

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

Proceeding as in the proof of Theorem 7.3, one shows that $C_{1,k}(\omega, T) = 0$ for $k > 2^N$, and (7.26) follows.

Let us now analyze $C_{1,k}(\omega, T)$ as $T \to \infty$. First note that, on the one hand, from (7.26) we have

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

On the other hand, we have

$$\mathbb{E}[y^{L_T^\omega(T-)} \mid L_T^\omega(0-) = 1] = \sum_{k=1}^\infty \mathbb{P}(L_T^\omega(T-) = k \mid L_T^\omega(0-) = 1)y^k.$$

Since $U^\top$ is the change-of-basis matrix from $(y^k)_{k \in \mathbb{N}}$ to $(S_k(y))_{k \in \mathbb{N}}$, we deduce that

$$C_{1,k}(\omega, T) = \sum_{i \in \mathbb{N}} u_{i,k}\mathbb{P}(L_T^\omega(T-) = i \mid L_T^\omega(0-) = 1) = \mathbb{E}[u_{L_T^\omega(T-),k} \mid L_T^\omega(0-) = 1]. \quad (7.33)$$

From Theorem 7.4, we know that the distribution of $L_T^\omega(T-)$ converges when $T \to \infty$. In addition, according to Lemma 7.5 the function $i \mapsto u_{i,k}$ is bounded, and hence $C_{1,k}(\omega, T)$

converges to a real number. Recall that $T_1 < T_2 < \cdots$ is the increasing sequence of the jump times of ω and that this sequence converges to infinity. Therefore

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

where we used (7.27) in the last step. This shows that the limit on the right-hand side of (7.29) exists and equals $\lim_{T \to \infty} C_{1,k}(\omega, T)$.

It remains to prove (7.28) together with the convergence of the corresponding series. We already know from Theorem 7.4 that $h_T^\omega(x)$ converges to $h^\omega(x)$ when $T \to \infty$, and we have just proved (7.26) and that for any $k \geq 1$, $C_{1,k}(\omega, T)$ converges to $C_{1,k}(\omega, \infty)$, defined in (7.29), when $T \to \infty$. Now we claim that, for all $y \in [0, 1]$ and $T > T_1$,

$$|C_{1,k}(\omega, T) S_k(y)| \leq 4^k \times (2ek)^{(k+\theta)/2} e^{-\lambda_k T_1}. \tag{Claim 6}$$

Assume that Claim 6 is true. Then (7.28) and the convergence of the series follow using the dominated convergence theorem. It only remains to prove Claim 6. As in the proof of (7.33), one shows that

$$\mathcal{P}_{(T-T_1)+}^T(\omega)\rho(y) = E[y^{L_T^\omega(T-T_1)} \mid L_T^\omega(0-) = 1] = \sum_{k=1}^\infty \tilde{C}_{1,k}(\omega, T) S_k(y),$$

where $\tilde{C}_{1,k}(\omega, T) = \mathbb{E}[u_{L_T^\omega(T-T_1),k} \mid L_T^\omega(0-) = 1]$. Proceeding as in the proof of (7.32), we can prove that

$$\begin{aligned} \mathcal{P}_{(T-T_1)+}^T(\omega)\rho(y) \\ = UE(T - T_N)\Phi(\Delta\omega(T_N))^\top E(T_N - T_{N-1})\Phi(\Delta\omega(T_{N-1}))^\top \cdots \Phi(\Delta\omega(T_1))^\top S(y). \end{aligned} \tag{7.34}$$

Since $E(T_1)$ is diagonal with entries $(e^{-\lambda_j T_1})_{j \in \mathbb{N}}$, we conclude from (7.32) and (7.34) that $C_{1,k}(\omega, T) = e^{-\lambda_k T_1} \tilde{C}_{1,k}(\omega, T)$. Therefore

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

This together with Lemma 7.5 (see below) implies that, for all $k \geq 1$ and $t \geq 0$,

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

Combining this with Lemma 7.7, we obtain Claim 6, which concludes the proof. $\qquad \square$

Lemma 7.5. *For all $k \geq 1$,*

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

Proof. Let $k \geq 1$. By the definition of the matrix U in Lemma 7.4, the sequence $(u_{j,k})_{j \geq 1}$ satisfies

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

$$u_{k+l,k} = \frac{1}{\lambda_{k+l} - \lambda_k} \left(\gamma_{k+l} u_{k+l-1,k} + \theta v_0 \sum_{j=0}^{l-2} u_{k+j,k} \right) \text{ for } l \geq 2.$$

Let $M_k^j := \sup_{i \leq j} u_{i,k}$. By the definitions of $\gamma_{j+1}, \lambda_{k+1}, \lambda_k$ (see Lemma 7.4), we have

$$\gamma_{k+1} = \lambda_{k+1} - (k-1)\theta v_0 > \lambda_{k+1} - \lambda_k.$$

This together with the above expressions yields that

$$M_k^k = 1, \qquad M_k^{k+1} = \frac{\lambda_{k+1} - (k-1)\theta v_0}{\lambda_{k+1} - \lambda_k} \leq \frac{\lambda_{k+1}}{\lambda_{k+1} - \lambda_k} = 1 + \frac{\lambda_k}{\lambda_{k+1} - \lambda_k}.$$

Moreover, for $l \geq 2$, we have

$$u_{k+l,k} \leq M_k^{k+l-1} \frac{\gamma_{k+l} + (l-1)\theta v_0}{\lambda_{k+l} - \lambda_k} = M_k^{k+l-1} \frac{\lambda_{k+l} - (k-1)\theta v_0}{\lambda_{k+l} - \lambda_k} \leq M_k^{k+l-1} \frac{\lambda_{k+l}}{\lambda_{k+l} - \lambda_k}.$$

Hence, we have

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

As a consequence, we have

$$\sup_{j \geq 1} u_{k,j} \leq \prod_{l=1}^{\infty} \left(1 + \frac{\lambda_k}{\lambda_{k+l} - \lambda_k} \right) =: M_k^{\infty}. \tag{7.35}$$

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

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

where we used Lemma 7.6 (see below) in the last step. Since $\lambda_k = (k-1)(k+\theta)$ (see Lemma 7.4), the desired result follows.   □

Lemma 7.6. *For all* $k \in \mathbb{N}$,

$$\sum_{l=1}^{\infty} \frac{1}{\lambda_{k+l} - \lambda_k} \leq \frac{\log(2ek)}{2(k-1)}.$$

Proof. Using the definition of λ_k in Lemma 7.4, we have

$$\sum_{l=1}^{\infty} \frac{1}{\lambda_{k+l} - \lambda_k} \leq \sum_{l=1}^{\infty} \frac{1}{(k+l)(k+l-1) - k(k-1)}$$

$$\leq \frac{1}{(k+1)k - k(k-1)} + \int_k^{\infty} \frac{1}{x(x+1) - k(k-1)} \mathrm{d}x$$

$$= \frac{1}{2k} + \int_1^{\infty} \frac{1}{u^2 + (2k-1)u} \mathrm{d}u = \frac{1}{2k} + \lim_{a \to \infty} \int_1^a \frac{1}{u(u+2k-1)} \mathrm{d}u$$

$$\leq \frac{1}{2k-1} \left[1 + \lim_{a \to \infty} \left(\int_1^a \frac{1}{u} \mathrm{d}u - \int_1^a \frac{1}{u+2k-1} \mathrm{d}u \right) \right]$$

$$= \frac{1}{2k-1} \left[1 + \lim_{a \to \infty} \log \left(\frac{a2k}{a+2k-1} \right) \right] \leq \frac{\log(2ek)}{2(k-1)}. \qquad \square$$

Lemma 7.7. *For all $k \in \mathbb{N}$, we have*

$$\sup_{y \in [0,1]} |S_k(y)| \leq 4^k.$$

Proof. By definition of the polynomials S_k in (7.23), we have for $k \geq 1$

$$\sup_{y \in [0,1]} |S_k(y)| \leq \sum_{i=1}^{k} |v_{k,i}|. \qquad (7.36)$$

Let us fix $k \geq 1$ and define $S_j^k := \sum_{i \geq j}^{k} |v_{k,i}|$. Note that $S_j^k = 0$ for $j > k$ and that $S_k^k = 1$ by the definition of the matrix $(v_{i,j})_{i,j \in \mathbb{N}}$ in Lemma 7.4. In particular, the result is true for $k = 1$. Thus, we assume that $k > 1$ from now on. Using (7.22), we see that for any $j \in [k-1]$,

$$S_j^k = S_{j+1}^k + |v_{k,j}| = S_{j+1}^k + \left| \frac{-1}{(\lambda_k - \lambda_j)} \left[\left(\sum_{l=j+2}^{k} v_{k,l} \right) \theta v_0 + v_{k,j+1} \gamma_{j+1} \right] \right|$$

$$\leq S_{j+1}^k + \frac{1}{(\lambda_k - \lambda_j)} \left[S_{j+2}^k \theta v_0 + (S_{j+1}^k - S_{j+2}^k) \gamma_{j+1} \right]$$

$$\leq \left(1 + \frac{\gamma_{j+1}}{\lambda_k - \lambda_j} \right) S_{j+1}^k + \frac{\theta v_0 - \gamma_{j+1}}{\lambda_k - \lambda_j} S_{j+2}^k.$$

Note that $\frac{\theta v_0 - \gamma_{j+1}}{\lambda_k - \lambda_j} \leq 0$, because of the definition of the coefficients γ_i in Lemma 7.4. Thus, for $j \in [k-1]$,

$$S_j^k \leq \left(1 + \frac{\gamma_{j+1}}{\lambda_k - \lambda_j} \right) S_{j+1}^k. \qquad (7.37)$$

By the definitions of $\gamma_{j+1}, \lambda_k, \lambda_j$ in Lemma 7.4, and using that $j < k$, we have

$$\frac{\gamma_{j+1}}{\lambda_k - \lambda_j} = \frac{(j+1)j + j\theta v_1 + \theta v_0}{k(k-1) - j(j-1) + (k-j)\theta v_1 + (k-j)\theta v_0}$$

$$\leq \frac{(j+1)j}{j(k-1) - j(j-1)} + \frac{j\theta v_1}{(k-j)\theta v_1} + \frac{\theta v_0}{(k-j)\theta v_0} \leq 2\frac{j+1}{k-j}.$$

In particular,

$$1 + \frac{\gamma_{j+1}}{\lambda_k - \lambda_j} \leq \frac{k+j+2}{k-j}.$$

Plugging this into (7.37) yields, for all $j \in [k-1]$,

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

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

$$\sum_{i=1}^{k} |v_{k,i}| = S_1^k \leq \prod_{j=1}^{k-1} \frac{k+j+2}{k-j} = \binom{2k+1}{k-1} = \frac{\binom{2k+1}{k-1} + \binom{2k+1}{k+2}}{2} \leq 2^{2k+1}/2 = 4^k.$$

Combining with (7.36), we obtain the desired result.   □

Acknowledgements

Funding information

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BAAKE, E., CORDERO, F. AND HUMMEL, S. (2022). Lines of descent in the deterministic mutation-selection model with pairwise interaction. *Ann. Appl. Prob.* **32**, 2400–2447.
[2] BAAKE, E., LENZ, U. AND WAKOLBINGER, A. (2016). The common ancestor type distribution of a Λ-Wright–Fisher process with selection and mutation. *Electron. Commun. Prob.* **21**, 16 pp.
[3] BAAKE, E. AND WAKOLBINGER, A. (2018). Lines of descent under selection. *J. Statist. Phys.* **172**, 156–174.
[4] BANSAYE, V., CABALLERO, M.-E. AND MÉLÉARD, S. (2019). Scaling limits of population and evolution processes in random environment. *Electron. J. Prob.* **24**, 38 pp.
[5] BANSAYE, V., KURTZ, T. G. AND SIMATOS, F. (2016). Tightness for processes with fixed points of discontinuities and applications in varying environment. *Electron. Commun. Prob.* **21**, 9 pp.
[6] BARCZY, M., LI, Z. AND PAP, G. (2015). Yamada–Watanabe results for stochastic differential equations with jumps. *Internat. J. Stoch. Anal.* **2015**, article no. 460472, 23 pp.
[7] BISWAS, N., ETHERIDGE, A. AND KLIMEK, A. (2021). The spatial Lambda-Fleming–Viot process with fluctuating selection. *Electron. J. Prob.* **26**, 51 pp.
[8] BULL, J. J. (1987). Evolution of phenotypic variance. *Evolution* **41**, 303–315.
[9] BÜRGER, R. AND GIMELFARB, A. (2002). Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genet. Res.* **80**, 31–46.

[10] CHETWYND-DIGGLE, J. AND KLIMEK, A. (2019). Rare mutations in the spatial Lambda-Fleming–Viot model in a fluctuating environment and SuperBrownian motion. Preprint. Available at https://arxiv.org/abs/1901.04374.

[11] CORDERO, F. (2017). Common ancestor type distribution: a Moran model and its deterministic limit. *Stoch. Process. Appl.* **127**, 590–621.

[12] CORDERO, F., HUMMEL, S. AND SCHERTZER, E. (2022). General selection models: Bernstein duality and minimal ancestral structures. *Ann. Appl. Prob.* **32**, 1499–1556.

[13] CORDERO, F. AND MÖHLE, M. (2019). On the stationary distribution of the block counting process for population models with mutation and selection. *J. Math. Anal. Appl.* **474**, 1049–1081.

[14] DYNKIN, E. B. (1965). *Markov Processes*, Vol. 1. Springer, Berlin, Heidelberg.

[15] ETHERIDGE, A. M., GRIFFITHS, R. C. AND TAYLOR, J. E. (2010). A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theoret. Pop. Biol.* **78**, 77–92.

[16] FOUCART, C. (2013). The impact of selection in the Λ-Wright–Fisher model. *Electron. Commun. Prob.* **18**, 10 pp.

[17] FU, Z. AND LI, Z. (2010). Stochastic equations of non-negative processes with jumps. *Stoch. Process. Appl.* **120**, 306–330.

[18] GILLESPIE, J. H. (1972). The effects of stochastic environments on allele frequencies in natural populations. *Theoret. Pop. Biol.* **3**, 241–248.

[19] GONZÁLEZ CASANOVA, A. AND SPANÒ, D. (2018). Duality and fixation in Ξ-Wright–Fisher processes with frequency-dependent selection. *Ann. Appl. Prob.* **28**, 250–284.

[20] GONZÁLEZ CASANOVA, A., SPANÒ, D. AND WILKE-BERENGUER, M. (2019). *The effective strength of selection in random environment*. Preprint. Available at https://arxiv.org/abs/1903.12121.

[21] GUILLIN, A., JABOT, F. AND PERSONNE, A. (2020). On the Simpson index for the Moran process with random selection and immigration. *Internat. J. Biomath.* **13**, article no. 2050046.

[22] GUILLIN, A., PERSONNE, A. AND STRICKLER, E. (2019). Persistence in the Moran model with random switching. Preprint. Available at https://arxiv.org/abs/1911.01108.

[23] JANSEN, S. AND KURT, N. (2014). On the notion(s) of duality for Markov processes. *Prob. Surveys* **11**, 59–120.

[24] KALLENBERG, O. (2021). *Foundations of Modern Probability*, 3rd edn. Springer, Cham.

[25] KARLIN, S. AND BENNY LEVIKSON, B. (1974). Temporal fluctuations in selection intensities: case of small population size. *Theoret. Pop. Biol.* **6**, 383–412.

[26] KARLIN, S. AND LIBERMAN, U. (1975). Random temporal variation in selection intensities: one-locus two-allele model. *J. Math. Biol.* **2**, 1–17.

[27] KARLIN, S. AND LIEBERMAN, U. (1974). Random temporal variation in selection intensities: Case of large population size. *Theoret. Pop. Biol.* **6**, 355–382.

[28] KIMURA, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719.

[29] KRONE, S. M. AND NEUHAUSER, C. (1997). Ancestral processes with selection. *Theoret. Pop. Biol.* **51**, 210–237.

[30] KURTZ, T. G. (2011). Equivalence of stochastic equations and martingale problems. In *Stochastic Analysis 2010*, ed. D. Crisan, Springer, Berlin, pp. 113–130.

[31] LENZ, U., KLUTH, S., BAAKE, E. AND WAKOLBINGER, A. (2015). Looking down in the ancestral selection graph: a probabilistic approach to the common ancestor type distribution. *Theoret. Pop. Biol.* **103**, 27–37.

[32] LI, Z. AND PU, F. (2012). Strong solutions of jump-type stochastic equations. *Electron. Commun. Prob.* **17**, 13 pp.

[33] MÖHLE, M. (2001). Forward and backward diffusion approximations for haploid exchangeable population models. *Stoch. Process. Appl.* **95**, 133–149.

[34] NEUHAUSER, C. (1999). The ancestral graph and gene genealogy under frequency-dependent selection. *Theoret. Pop. Biol.* **56**, 203–214.

[35] NEUHAUSER, C. AND KRONE, S. M. (1997). The genealogy of samples in models with selection. *Genetics* **145**, 519–534.

[36] SAGITOV, S., JAGERS, P. AND VATUTIN, V. (2010). Coalescent approximation for structured populations in a stationary random environment. *Theoret. Pop. Biol.* **78**, 192–199.

[37] VAN CASTEREN, J. A. (1992). On martingales and Feller semigroups. *Results Math.* **21**, 274–288.

Appendix A. J_1-Skorokhod topology and weak convergence

A.1 Definitions and remarks on the J_1-Skorokhod topology

For $T > 0$, as in the beginning of Section 2 we denote by $\mathbb{D}_{0,T}$ the space of càdlàg functions in $[0, T]$ with values on $\mathbb{R}$. Let $\mathcal{C}_T^{\uparrow}$ denote the set of increasing, continuous functions from

[0, T] onto itself. For $\lambda \in \mathcal{C}_T^\uparrow$, we set

$$\|\lambda\|_T^0 := \sup_{0 \le u < s \le T} \left| \log \left(\frac{\lambda(s) - \lambda(u)}{s - u} \right) \right|. \tag{A.1}$$

We define the Billingsley metric d_T^0 in $\mathbb{D}_{0,T}$ via

$$d_T^0(f, g) := \inf_{\lambda \in \mathcal{C}_T^\uparrow} \{ \|\lambda\|_T^0 \vee \|f - g \circ \lambda\|_{T,\infty} \}, \quad \text{where } \|f\|_{T,\infty} := \sup_{s \in [0,T]} |f(s)|. \tag{A.2}$$

The metric d_T^0 induces the J_1-Skorokhod topology in $\mathbb{D}_{0,T}$. An important feature is that the space $(\mathbb{D}_{0,T}, d_T^0)$ is separable and complete. The role of the time-change λ in the definition of d_T^0 is to capture the fact that two càdlàg functions can be close in spite of a small difference between their jumping times.

For $T > 0$, a function $\omega \in \mathbb{D}_{0,T}$ is said to be pure-jump if $\sum_{u \in (0,T]} |\Delta\omega(u)| < \infty$ and for all $t \in (0, T]$,

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

where $\Delta\omega(u) := \omega(u) - \omega(u-)$, $u \in [0, T]$. In the set of pure-jump functions, we consider the following metric:

$$d_T^\star(\omega_1, \omega_2) := \inf_{\lambda \in \mathcal{C}_T^\uparrow} \left\{ \|\lambda\|_T^0 \vee \sum_{u \in [0,T]} |\Delta\omega_1(u) - \Delta(\omega_2 \circ \lambda)(u)| \right\}. \tag{A.3}$$

The next result provides comparison inequalities between the metrics d_T^0 and $d_T^\star$.

Lemma A.1 *Let ω_1 and ω_2 be two pure-jump functions with $\omega_1(0) = \omega_2(0) = 0$; then*

$$d_T^0(\omega_1, \omega_2) \le d_T^\star(\omega_1, \omega_2).$$

If ω_1 and ω_2 are non-decreasing, and ω_1 jumps exactly n times in $[0, T]$, then

$$d_T^\star(\omega_1, \omega_2) \le (4n + 3) d_T^0(\omega_1, \omega_2).$$

Proof. Let $\lambda \in \mathcal{C}_T^\uparrow$ 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 $t \in [0, T]$,

$$|f(t) - g(t)| = \left| \sum_{u \in [0,t]} (\Delta f(u) - \Delta g(u)) \right| \le \sum_{u \in [0,t]} |\Delta f(u) - \Delta g(u)|.$$

The first inequality follows. Now, assume that ω_1 and ω_2 are non-decreasing and that ω_1 has n jumps in $[0, T]$. Let $t_1 < \cdots < t_n$ be the consecutive jump times of ω_1. We first prove that, for any $k \in [n]$,

$$\sum_{u \in [0,t_k]} |\Delta f(u) - \Delta g(u)| \le (4k + 1) \|f - g\|_{t_k, \infty}, \tag{A.4}$$

where $\|\cdot\|_{t,\infty}$, $t > 0$, is defined in (A.2). We proceed by induction on k. Note that

$$\sum_{u \in [0,t_1]} |\Delta f(u) - \Delta g(u)| = \sum_{u \in [0,t_1)} \Delta g(u) + |\Delta f(t_1) - \Delta g(t_1)| \le g(t_1 -) + 2\|f - g\|_{t_1,\infty}$$
$$\le 3\|f - g\|_{t_1,\infty},$$

which proves (A.4) for $k = 1$. Now, assuming that (A.4) is true for $k \in [n-1]$, we obtain

$$\sum_{u \in [0,t_{k+1}]} |\Delta f(u) - \Delta g(u)| = \sum_{u \in [0,t_k]} |\Delta f(u) - \Delta g(u)| + \sum_{u \in (t_k,t_{k+1})} \Delta g(u)$$
$$+ |\Delta f(t_{k+1}) - \Delta g(t_{k+1})|$$
$$\le (4k+1)\|f - g\|_{t_k,\infty} + g(t_{k+1} -) - g(t_k) + 2\|f - g\|_{t_{k+1},\infty}$$
$$= (4k+1)\|f - g\|_{t_k,\infty} + (g(t_{k+1} -) - f(t_{k+1} -)) - (g(t_k) - f(t_k)) + 2\|f - g\|_{t_{k+1},\infty}$$
$$\le (4k+1)\|f - g\|_{t_k,\infty} + 4\|f - g\|_{t_{k+1},\infty} \le (4(k+1)+1)\|f - g\|_{t_{k+1},\infty}.$$

Hence, (A.4) also holds for $k + 1$. This ends the proof of (A.4) by induction. Finally, using (A.4), we get

$$\sum_{u \in [0,T]} |\Delta f(u) - \Delta g(u)| = \sum_{u \in [0,t_n]} |\Delta f(u) - \Delta g(u)| + \sum_{u \in (t_n,T]} \Delta g(u)$$
$$\le (4n+1)\|f - g\|_{t_n,\infty} + g(T) - g(t_n) \le (4n+3)\|f - g\|_{T,\infty},$$

ending the proof. □

A.2. Bounded Lipschitz metric and weak convergence

Let (E, d) denote 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, whose definition is recalled in this section.

Definition 1. *(Lipschitz function.)* A real-valued function F on (E, d) is said to be Lipschitz if there is $K > 0$ such that

$$|F(x) - F(y)| \le K d(x, y), \quad \text{for all } x, y \in E.$$

We denote by $\mathrm{BL}(E)$ the vector space of bounded Lipschitz functions on E. The space $BL(E)$ is equipped with the norm

$$\|F\|_{\mathrm{BL}} := \sup_{x \in E} |F(x)| \vee \sup_{x,y \in E: x \ne y} \left\{ \frac{|F(x) - F(y)|}{d(x, y)} \right\}, \quad F \in BL(E). \tag{A.5}$$

Definition 2. *(Bounded Lipschitz metric.)* Let μ, ν be two probability measures on E. The bounded Lipschitz distance between μ and ν is defined by

$$\varrho_E(\mu, \nu) := \sup \left\{ \left| \int F d\mu - \int F d\nu \right| : F \in \mathrm{BL}(E), \ \|F\|_{\mathrm{BL}} \le 1 \right\}. \tag{A.6}$$

The bounded Lipschitz distance defines 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.

$$\varrho_E(\mu_n, \mu) \xrightarrow[n\to\infty]{} 0 \quad \Longleftrightarrow \quad \mu_n \xrightarrow[n\to\infty]{(d)} \mu.$$

Appendix B. Table of notation

Notation	Meaning	First appears
$X = (X(t))_{t\geq 0}$	Solution of (1.3)	Sec. 1
σ, θ, v_0, v_1	Parameters of selection and mutation	Sec. 1
$J = (J(t))_{t\geq 0}$	Pure-jump subordinator	Sec. 1
μ	Lévy measure of J	Sec. 1
$\zeta = (t_i, p_i)_{i\in I}$	Collection of jumps of J	Sec. 1
$S(t)$	$\sigma t + J(t)$	Sec. 1
$\mathbb{D}_{s,t}, \mathbb{D}$	Spaces of càdlàg functions	Sec. 2
$\omega = (\omega(t))_{t\geq 0}$	Deterministic environment	Sec. 2.1
$\mathbb{D}_T^\star/\mathbb{D}^\star$	Set of environments $(\omega(t))_{t\in[0,T]}/\omega$	Sec. 2.1
σ_N, θ_N	Parameters of Moran model	Sec. 2.1
$\Delta\omega(t)$	Jump of ω at time t	Sec. 2.1
$\mathbf{0}$	Null environment	Sec. 2.1
$(Z_N^\omega(t))_{t\geq 0}$	Quenched $\sharp$ of fit individuals	Sec. 2.1
$(Z_N^J(t))_{t\geq 0}$	Annealed $\sharp$ of fit individuals	Sec. 2.2
$\mathcal{A}_N$	Generator of $(Z_N^J(t))_{t\geq 0}$	Sec. 2.2
$\mathcal{A}_N^0$	$\mathcal{A}_N$ in the case $\mu = 0$	Sec. 2.2
$X^\omega = (X^\omega(t))_{t\geq 0}$	Quenched version of X	Sec. 2.3
J_N, ω_N	$J(\cdot/N), \omega(\cdot/N)$	Sec. 2.3
X_N, X_N^ω	$Z_N^{J_N}(N\cdot)/N, Z_N^{\omega_N}(N\cdot)/N$	Sec. 2.3
$\bar{J}^T = (\bar{J}^T(\beta))_{\beta\in[0,T]}$	Time reversal of J	Sec. 2.4
$(\mathcal{G}(\beta))_{\beta\geq 0}/(\mathcal{G}_T^\omega(\beta))_{\beta\geq 0}$	Annealed/quenched ASG	Sec. 2.4
$\sigma_{m,k}$	Rates of simultaneous branchings	Sec. 2.4
$(\bar{\mathcal{G}}(\beta))_{\beta\geq 0}/(\bar{\mathcal{G}}_T^\omega(\beta))_{\beta\geq 0}$	Annealed/quenched killed ASG	Sec. 2.5
$(R(\beta))_{\beta\geq 0}/(R_T^\omega(\beta))_{\beta\geq 0}$	Line-counting process of $\bar{\mathcal{G}}/\bar{\mathcal{G}}_T^\omega$	Sec. 2.5
$Q_\dagger^\mu = (q_\dagger^\mu(i,j))_{i,j\in\mathbb{N}_0^\dagger}$	Generator of $(R(\beta))_{\beta\geq 0}$	Sec. 2.5
$\pi_n/\Pi_n(\omega)$	Absorption probabilities of R/R_T^ω	Sec. 2.5
$\eta_X/\mathcal{L}^\omega$	Limit distribution of $X(t)/X^\omega(0)$	Sec. 2.5
$X(\infty)$	Random variable with law η_X	Sec. 2.5
$J \otimes_\tau \omega$	Mixed environment	Sec. 2.5
$h_T(x)/h_T^\omega(x)$	Ancestral type distribution at T	Sec. 2.6
$h(x)/h^\omega(x)$	Ancestral type distribution	Sec. 2.6

Notation	Meaning	First appears
$(L(\beta))_{\beta \geq 0}/(L_T^\omega(\beta))_{\beta \geq 0}$	pLD-ASG's line-counting process	Sec. 2.6
$Q^\mu = (q^\mu(i,j))_{i,j \in \mathbb{N}}$	Generator of $(L(\beta))_{\beta \geq 0}$	Sec. 2.6
η_L/μ^ω	Limit law of $L(T)/L_T^\omega(T-)$	Sec. 2.6
$L(\infty)$	Random variable with law η_L	Sec. 2.6
a_n	$\mathbb{P}(L(\infty) > n)$	Sec. 2.6
$\gamma_{i,j}$	Coefficients in recursion (2.18)	Sec. 2.6
$\mu_N(\omega)$	Law of $(Z_N^\omega(t))_{t \in [0,T]}$	Sec. 3.1
$\omega^\delta/\omega_\delta$	ω with small/large jumps removed	Sec. 3.1
$\mathcal{A}_N^*/\mathcal{A}$	Generator of X_N/X	Sec. 3.2
E_N	State space of X_N	Sec. 3.2
$d_T^0/d_T^\star$	Metric on $\mathbb{D}_{0,T}/\mathbb{D}_T^\star$	App. A.1
$\mathcal{C}_T^\uparrow$	Increasing, continuous $f : [0,T] \to [0,T]$	App. A.1
$\|\cdot\|_T^0, \|\cdot\|_{T,\infty}$	Functional norms	App. A.1
$BL(E)$	Bounded Lipschitz functions $F : E \to \mathbb{R}$	App. A.2
$\|\cdot\|_{\mathrm{BL}}$	Norm on $BL(E)$	App. A.2
$\varrho_E(\cdot, \cdot)$	Bounded Lipschitz metric	App. A.2