Electronic Journal of Probability

Electron. J. Probab. **17** (2012), no. 31, 1–36.
ISSN: 1083-6489 DOI: 10.1214/EJP.v17-1954

Fixation probability for competing selective sweeps*

Charles Cuthbertson† Alison Etheridge‡ Feng Yu§

Abstract

We consider a biological population in which a beneficial mutation is undergoing a selective sweep when a second beneficial mutation arises at a linked locus. We investigate the probability that both mutations will eventually fix in the population. Previous work has dealt with the case where the second mutation to arise confers a smaller benefit than the first. In that case population size plays almost no rôle. Here we consider the opposite case and observe that, by contrast, the probability of both mutations fixing can be heavily dependent on population size. Indeed the key parameter is rN, the product of the population size and the recombination rate between the two selected loci. If rN is small, the probability that both mutations fix can be reduced through interference to almost zero while for large rN the mutations barely influence one another. The main rigorous result is a method for calculating the fixation probability of a double mutant in the large population limit.

Keywords: selective sweep; fixation probability; double mutant.
AMS MSC 2010: Primary 92D15, Secondary 82C22; 60J05; 92D10.
Submitted to EJP on May 25, 2010, final version accepted on March 11, 2012.

1 Introduction

Natural populations incorporate beneficial mutations through a combination of chance and the action of natural selection. The process whereby a beneficial mutation arises (in what is generally assumed to be a large and otherwise neutral population) and eventually spreads to the entire population is called a *selective sweep*. When beneficial mutations are rare, we can make the simplifying assumption that selective sweeps do not overlap. A great deal is known about such isolated selective sweeps (see e.g. Chapter 5 of Ewens 2004). Haldane (1927) showed that under a discrete generation haploid model, the probability that a beneficial allele with selective advantage σ eventually *fixes* in a diploid population of size $2N$, i.e. its frequency increases from $1/(2N)$ to 1, is approximately 2σ. Much less is understood when selective sweeps overlap, that is when further beneficial mutations arise at different loci during the timecourse of a sweep. Our aim here is to investigate the impact of the resulting interference in the case when

*CC supported by EPSRC DTA, FY supported by EPSRC/I028498/1.
†Morgan Stanley, New York, USA. E-mail: charles.cuthbertson@gmail.com
‡University of Oxford, UK. E-mail: etheridg@stats.ox.ac.uk
§University of Bristol, UK. E-mail: feng.yu@bristol.ac.uk

two sweeps overlap. In particular, we shall investigate the probability that both beneficial mutations eventually become fixed in the population. In reality, beneficial mutations may arise at a third (or fourth, fifth, etc.) locus during a selective sweep. Therefore, this work can be regarded as a first step towards understanding the far more challenging problem of multiple overlapping sweeps.

Before stating our model and results more precisely let us recall some key concepts and place our work in context. Because genes are organised on chromosomes and chromosomes are in turn grouped into individuals, different genetic loci do not evolve independently of one another. In a diploid population (in which chromosomes are carried in pairs), however, nor are chromosomes passed down as intact units. A given chromosome is inherited from one of the two parents, but *recombination* or *crossover* events can result in the allelic types at two distinct loci being inherited one from each of the corresponding pair of chromosomes in the parent. We refer to these chromosomes as 'individuals'. Fisher (1930) and Muller (1932) were the first to give verbal arguments for the evolutionary advantage of such crossover events. Hill & Robertson (1966) provide the first theoretical reasoning, which differs from the arguments of Fisher and Muller, but is mathematically equivalent. They support their arguments with simulations. The basic idea is that selectively beneficial alleles occurring on linked loci interfere with each other. In the absence of recombination, they can only both become fixed in the population if one arises on a chromosome which already carries the other. As a result, the probability of fixation of beneficial alleles is reduced. Through recombination, alleles at two linked sites on the same chromosome can become fixed in a population even if they are not initially associated. This has come to be known as the Hill-Robertson effect or the Fisher-Muller effect.

Each individual in the population will have a type denoted ij where $i, j \in \{0, 1\}$. We use the first and second digit, respectively, to indicate whether or not the individual carries the older or more recent beneficial mutation, and assume that the fitness effects of these two mutations are additive. We consider first the frequency of the older mutation until the time when the second mutation arises. During this period, it is undergoing an isolated selective sweep. Suppose that a single advantageous allele with selective advantage σ_1 arises in an otherwise neutral (type 00) population of size $2N$, corresponding to a diploid population of size N. We use X_{ij} to denote the proportion of individuals of type ij. If $2N\sigma_1$ is large, then the frequency of the favoured allele, X_{10}, will be well-approximated by the solution to the stochastic differential equation

$$dX_{10} = \sigma_1 X_{10}(1 - X_{10})\, dt + \sqrt{\frac{1}{2N} X_{10}(1 - X_{10})}\, dW, \tag{1.1}$$

where $\{W(t)\}_{t \geq 0}$ is a standard Wiener process (Ethier & Kurtz 1986, Eq. 10.2.7[1]). If the favoured allele reaches frequency p, then the probability that it ultimately fixes is

$$\frac{1 - e^{-4N\sigma_1 p}}{1 - e^{-4N\sigma_1}}. \tag{1.2}$$

If we assume $2N\sigma_1$ to be large and $p = 1/(2N)$, then the above is approximately $1 - e^{-2\sigma_1}$. If a sweep *does* take place then (conditional on fixation) we obtain

$$d\tilde{X}_{10} = \sigma_1 \tilde{X}_{10}(1 - \tilde{X}_{10}) \coth(N\sigma_1 \tilde{X}_{10})\, dt + \sqrt{\frac{1}{2N} \tilde{X}_{10}(1 - \tilde{X}_{10})}\, dW \tag{1.3}$$

[1] Note that the process described in this equation is in the diffusion time scale, i.e. one unit of time equals roughly $2N$ generations. In our case, we slow down the diffusion time scale by a factor of $2N$, so that the time units of (1.1) are generations.

and from this it is easy to calculate the expected duration of the sweep. Writing $\tilde{T}_{fix} = \inf\{t \geq 0 : \tilde{X}_{10}(t) = 1 \Big| \tilde{X}_{10}(0) = 1/(2N)\}$, we have via a Green's function calculation (see for example Karlin & Taylor 1981) that

$$\mathbb{E}[\tilde{T}_{fix}] = \frac{2}{\sigma_1}\log(2N\sigma_1) + \mathcal{O}(1) \tag{1.4}$$

and the variance $var[\tilde{T}_{fix}]$ is $\mathcal{O}(1)$ as was noted, for example, by Kimura & Ohta (1969) and Etheridge, Pfaffelhuber & Wakolbinger (2006). More generally, an analogous Green function calculation to that leading to equation (1.4) gives that the expected time for the selected locus to reach frequency $\epsilon(N)$ is $\log(2N\sigma_1\epsilon(N))/\sigma_1 + \mathcal{O}(1)$. This is the same as the expected time for $\tilde{X}_{10}$ to increase from $1-\epsilon(N)$ to 1. On the other hand, for $\delta = \mathcal{O}(1)$, i.e. if δ is between $1/C$ and C for some constant C as $N \to \infty$ all other parameters being constant, the time for $\tilde{X}_{10}$ to increase from δ to $1-\delta$ is $\mathcal{O}(1)$. As a result, for large populations, during almost all of the timecourse of the sweep $\tilde{X}_{10}$ is either close to zero or close to one.

Now suppose that during the selective sweep of type 10 described by (1.1), more specifically, when X_{10} reaches a level U, another beneficial mutation with selection coefficient σ_2 occurs at a second linked locus in a randomly chosen individual. The recombination rate between these two loci is r, i.e. the expected number of recombination events per individual per generation between these two loci during a short time period of length Δt is approximately $r\Delta t$. We treat the arrival time of the second mutation as being uniformly distributed over the timecourse of the sweep of the first mutation. If N is large, then we can expect either U or $1-U$ to be close to 0 but $\gg 1/(2N)$. We work under this assumption in the remainder of this article.

Throughout this work, we will use the non-rigorous notion of a type becoming 'established'. By this we mean that the number of individuals of that type is much larger than 1, so that the subsequent trajectory of its frequency is roughly deterministic, at least until it again approaches either 0 or 1. Note that for the fittest type in the population (i.e. type 10 if the second beneficial mutation never arises or type 11 if there are two beneficial mutations) the difference between the establishment probability and the fixation probability can be bounded above by the expression in (1.2), with σ_1 taken to to be fitness difference between the fittest and second fittest types. The difference between the two probabilities is therefore very small if N is large.

The second mutation can arise in either a type 00 or a type 10 individual, forming a single type 01 individual in the former case, and a 11 individual in the latter case. If the second mutation arises during the first half (in terms of time) of the sweep of the first mutation, then U is likely to be very small and it is more likely for a type 01 individual to be formed. Otherwise, the second mutation arises during the second half of the sweep and the formation of a type 11 individual is more likely.

The case of the second beneficial mutation forming a type 11 individual is relatively straightforward. We assume the type 11 individual arises in a population consisting of $2NU$ type 10 and $2N(1-U)$ type 00 individuals. For very large N, U is likely to be close to either 0 or 1. Since type 11 is fitter than all other types, its fixation is almost certain once it becomes 'established' in the population. For large N, it only takes a short time to determine whether type 11 establishes itself, and we can assume the proportion of type 10 individuals remains roughly constant during this time since it is likely to be close to either 0 or 1 when type 11 arises. Hence the fixation probability of type 11 is essentially its establishment probability, which is approximately $2(\sigma_2 + \sigma_1(1-U))$, twice the 'effective' selective advantage of type 11 in a population consisting of $2NU$ type 10 and $2N(1-U)$ type 00 individuals. Hence we arrive at the following fact: if the second beneficial mutation arises in one of $2NU$ type 10 individuals, and σ_1 and σ_2 are both fixed, then the fixation probability of type 11 approaches

$$2(\sigma_2 + \sigma_1(1 - U)) \tag{1.5}$$

as $N \to \infty$. By 'fact', we mean this statement is easy to establish under a reasonable model, e.g. the one we give in §2.1.

The case of the second beneficial mutation forming a type 01 individual is far more interesting. In order for both mutations to sweep through the population, recombination must produce an individual carrying both mutations. The relative strength of selection acting on the two loci now becomes important. The case of $\sigma_1 > \sigma_2$ has been dealt with by Barton (1995) and Otto & Barton (1997), where they showed that the probability of fixation of the allele carrying both beneficial mutations depends on just two parameters, the ratio σ_2/σ_1 and the scaled recombination rate r/σ_1. Since type 10 is already present in significant numbers when the new mutation arises (and type 10 is fitter than type 01), the trajectory of X_{10} is well approximated by the logistic growth curve $1/(1 + \exp(-\sigma_1 t))$ until X_{11} reaches a level of $\mathcal{O}(1)$. At that point, fixation of type 11 is all but certain. Barton (1995) uses a branching process approximation to estimate the establishment probability of a type 11 individual produced by recombination. In particular, his approach is *independent* of population size. Not surprisingly, he finds that the fixation probability of the second mutation is reduced if it arises as a type 01 individual, but increased if it arises as a type 11 individual. Simulation studies performed in Otto & Barton (1997) confirm these findings in the case $\sigma_1 > \sigma_2$.

McVean & Charlesworth (2000) considered weakly selected mutations on two loci and argued that the Hill-Robertson effect is an important force in the evolution of non-recombining genomes. This corresponds to choosing σ_1 and σ_2 to be $\mathcal{O}(1/N)$. In contrast, we consider the case of strongly selected mutations such that σ_1 and σ_2 are $\mathcal{O}(1)$.

Gillespie (2001) considers the effects of repeated substitutions at a strongly selected locus on a completely linked (i.e. there is no recombination) *weakly selected* locus, extending his work in Gillespie (2000), where he considers a linked *neutral* locus. He too sees little dependence of his results on population size, leading him to suggest repeated genetic hitchhiking events as an explanation for the apparent insensitivity of the genetic diversity of a population to its size. Kim (2006) extends the work of Gillespie (2001) by considering the effect of repeated sweeps on a tightly (but not completely) linked locus. This whole body of work is concerned, in our terminology, with $\sigma_1 > \sigma_2$.

The case of $\sigma_2 > \sigma_1$ brings quite a different picture. The analysis used in Barton (1995) breaks down for the following reason: if type 01 gets established in the population then, since the second beneficial mutation is more competitive than the first, it is destined to start a sweep itself. Once X_{01} reaches $\mathcal{O}(1)$, X_{10} is no longer well approximated by a logistic growth curve and in fact will decrease to 0. The fixation probability of type 11 will then depend on the nonlinear interaction of all four types, $\{11, 01, 10, 00\}$, and our analysis will show that it is *heavily dependent* on population size. See Figure 1 below.

This paper is organized as follows. In §2.1 we set up a continuous time Moran model for the evolution of our population. In the biological literature, it would be more usual to consider a Wright-Fisher model, in which the population evolves in discrete, non-overlapping generations. The choice of a Moran model, in which generations overlap, is a matter of mathematical convenience. One expects similar results for a Wright-Fisher model. The choice of a discrete individual based model rather than a diffusion is forced upon us by our method of proof, but is anyway natural in a setting where population size plays a rôle in the results. A brief discussion of our model, for very large N, in §2 leads to a method for calculating the asymptotic ($N \to \infty$) fixation probability of type 11 when $\sigma_2 > \sigma_1$ and when the arrival of the second beneficial mutation forms a type

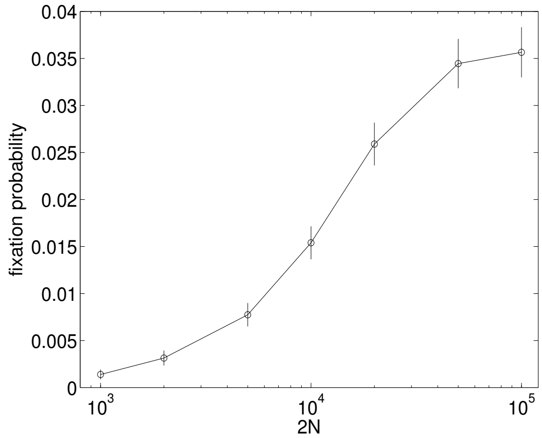


Figure 1: Simulation results for fixation probability of type 11 for the following initial condition: the second mutation arises in a type 00 individual, when $\lfloor (2N)^{0.7} \rfloor$ individuals in the population have the first mutation (i.e. are of type 10). Vertical bars denote two standard deviations. For an explanation of why $(2N)^{1-\zeta}$ for $\zeta \in (0,1)$ is the right initial condition, see the discussion before (2.3). Parameter values: $\sigma_1 = 0.012$, $\sigma_2 = 0.02$, $r = 4 \times 10^{-5}$ (recombination coefficient).

01 individual. Our main result, Theorem 2.3, is concerned with the case when the proportion U of type 10 individuals is equal to $(2N)^{-\zeta}$, where $\zeta < \sigma_1/\sigma_2 < 1$, at the time of the creation of the second beneficial mutation. We will see that in order for a non-trivial result to be obtained, i.e. for the effect of the interference to be intermediate between complete linkage (that is, no recombination) and independence, the recombination rate r must be $\mathcal{O}(1/N)$. The significance of the condition $\zeta < \sigma_1/\sigma_2$ will be explained in §2.2, and the complementary case of $\sigma_1/\sigma_2 < \zeta < 1$ will be discussed in more detail in §2.3, where we also discuss the case of moderate N. The rest of the paper is devoted to proofs, with §3 containing the proof of Theorem 2.3 and §4 containing the proof of Proposition 3.1. Results in §4 rely on supporting lemmas of §5.

2 Model and Results

2.1 A Moran Model for Two Competing Selective Sweeps

In this section we describe our model for the evolution of two competing selective sweeps. We use the notation from the introduction for the four possible types of individuals in the population $I = \{00, 01, 10, 11\}$, and assume that at the time when the second mutation arises, the proportion $U \in \{0, 1, \ldots, 2N\}/(2N)$ of type 10 individuals in the population is known. From now on we use $t = 0$ to denote the time when the second mutation arises. As explained in §1, since we are interested in the case when the time of arrival of the second beneficial mutation is uniformly distributed over the timecourse of the selective sweep of the first, we may assume that $2NU$ is much larger than 1.

For simplicity, we use an additive fitness model, even though an extension to more general cases should be straightforward, as long as type 11 is the fittest of all 4 types. Let $\sigma \in [0,1]$ be the selective advantage of the second beneficial mutation and $\sigma\gamma$ be the selective advantage of the first beneficial mutation (for some $\gamma \in (0,1)$). This corresponds to taking $\sigma_2 = \sigma$ and $\sigma_1 = \sigma\gamma$ in the notation of the introduction. The recombination rate between the two selected loci is denoted by r which we assume to be $o(1)$. If r and σ are small, then decoupling recombination from the rest of the reproduction process does not affect the behaviour of the model a great deal and it will simplify analysis. We use $\{\eta_n\zeta_n, n = 1, \ldots, 2N\}$ to denote the types of individuals in the population. At time $t = 0$, we assume that the population of $2N$ individuals consists of $2N(1-U)-1$ type 00 individuals, $2NU$ type 10 individuals and 1 type 01 individual. The dynamics of the model are as follows:

1. *Recombination*: Each ordered pair of individuals, $\eta_m\zeta_m$ and $\eta_n\zeta_n \in I$, is chosen at rate $r/(2N)$. With probability $1/2$, $\eta_m\zeta_n$ replaces $\eta_m\zeta_m$. Otherwise, $\eta_n\zeta_m$ replaces $\eta_m\zeta_m$.
2. *Resampling (and selection)*: Each ordered pair of individuals, $\eta_m\zeta_m$ and $\eta_n\zeta_n \in I$, is chosen at rate $1/(2N)$. With probability $p(\eta_m\zeta_m, \eta_n\zeta_n)$ given by
$$p(ij, kl) := \frac{1}{2}(1 + \sigma\gamma(i-k) + \sigma(j-l)),$$
a type $\eta_m\zeta_m$ individual replaces $\eta_n\zeta_n$. Otherwise a type $\eta_n\zeta_n$ individual replaces $\eta_m\zeta_m$.

Remark 2.1. *Evidently we must assume $\sigma > 0$ and $\sigma(1+\gamma) \leq 1$ to ensure that all probabilities used in the definition of the model are in $[0,1]$.*

Remark 2.2. *The recombination mechanism we use above actually corresponds to 'gene conversion'. In order to model crossover events one replaces mechanism 1 by the following:*

1'. Each individual is picked at rate r and replaced by one of type ij with probability $X_{i\bullet}X_{\bullet j}$, where $X_{i\bullet} = X_{i0} + X_{i1}$ and $X_{\bullet j} = X_{0j} + X_{1j}$.

If we use mechanism 1' and 2 as our model, the jump rates of X_{ij} will be slightly different from those in (2.1). The drift terms for X_{ij} (i.e. $r^+_{ij} - r^-_{ij}$ below), however, remain the same for both models. When X_{11} is small, the dominant contribution to the rate at which types 01 and 10 recombine to produce type 11 individuals is $2rX_{01}X_{10}$ for both models. As a result, both Theorem 2.3 and its proof apply to the model with mechanisms 1' and 2. We use the gene conversion model in Theorem 2.3 and its proof as its jump rates are a bit easier to write down.

Let $\mathbb{P}$ denote the law of this Moran particle system, and r^+_{ij} and r^-_{ij} be the rates at which X_{ij} increases and decreases by $1/(2N)$, respectively, then

$$\begin{aligned}
r^+_{01} &= NX_{01}[(1+\sigma)(1-X_{01}) - \sigma(1+\gamma)X_{11} - \sigma\gamma X_{10}] \\
&\qquad + rN(2X_{11}X_{00} + X_{01}X_{11} + X_{01}X_{00}) \\
r^-_{01} &= NX_{01}[(1-\sigma)(1-X_{01}) + \sigma(1+\gamma)X_{11} + \sigma\gamma X_{10}] \\
&\qquad + rNX_{01}(X_{00} + 2X_{10} + X_{11}) \\
r^+_{10} &= NX_{10}[(1+\sigma\gamma)(1-X_{10}) - \sigma(1+\gamma)X_{11} - \sigma X_{01}] \\
&\qquad + rN(X_{00}X_{10} + X_{11}X_{10} + 2X_{11}X_{00}) \\
r^-_{10} &= NX_{10}[(1-\sigma\gamma)(1-X_{10}) + \sigma(1+\gamma)X_{11} + \sigma X_{01}] \\
&\qquad + rNX_{10}(X_{00} + 2X_{01} + X_{11}) \\
r^+_{11} &= NX_{11}[(1+\sigma(1+\gamma))(1-X_{11}) - \sigma X_{01} - \sigma\gamma X_{10}] \\
&\qquad + rN(2X_{01}X_{10} + X_{01}X_{11} + X_{10}X_{11}) \\
r^-_{11} &= NX_{11}[(1-\sigma(1+\gamma))(1-X_{11}) + \sigma X_{01} + \sigma\gamma X_{10}] \\
&\qquad + rNX_{11}(2X_{00} + X_{10} + X_{01})
\end{aligned}$$

$$\begin{aligned} r_{00}^{+} &= NX_{00}[1 - X_{00} - \sigma(1+\gamma)X_{11} - \sigma X_{01} - \sigma\gamma X_{10}] \\ &\qquad +rN(X_{10}X_{00} + X_{00}X_{01} + 2X_{10}X_{01}) \\ r_{00}^{-} &= NX_{00}[1 - X_{00} + \sigma(1+\gamma)X_{11} + \sigma X_{01} + \sigma\gamma X_{10}] \\ &\qquad +rNX_{00}(X_{10} + 2X_{11} + X_{01}). \end{aligned} \tag{2.1}$$

To understand these rates a little better, consider for example the term r_{01}^{+}. The part due to selection is $NX_{01}[(1+\sigma)(1-X_{01}) - \sigma(1+\gamma)X_{11} - \sigma\gamma X_{10}]$. It must take account of the fitness of type 01 relative to the rest of the population. Thus, using the fitness of type 00 as a baseline, not only does it encode the advantage of type 01 over type 00 through the term $\sigma(1-X_{01})$, but also of the advantage of types 11 and 10 over 00 through the two negative terms.

We define the fixation time of this Moran particle system

$$T_{fix} = \inf\{t \geq 0 : X_{ij}(t) = 1 \text{ for some } ij \in I\}.$$

We observe that the Markov chain (X_{00}, X_{10}, X_{01}) has finitely many states and the recurrent states are $R = \{(0,0,0), (0,0,1), (0,1,0), (1,0,0)\}$. Every other state is transient and there is positive probability of reaching R starting from any transient state in finite time. Therefore $\mathbb{E}[T_{fix}] < \infty$, and in particular,

$$T_{fix} < \infty \ a.s.$$

2.2 Main result

We are concerned primarily with the case of very large population sizes, which is the regime where our main rigorous result, Theorem 2.3, operates. A non-rigorous analysis for moderate population sizes based on very similar ideas is also possible and appears in Yu & Etheridge (2010). In this section we focus on the case when the arrival of the new mutation results in a type 01 individual. To motivate our result, we consider the hurdles that must be overcome if type 11 is to become fixed in the population. Our approach will be to estimate the probability that each of these hurdles is overcome. First, following the appearance of the new mutant, X_{01} must 'become established', by which we mean achieve appreciable numbers in the population, e.g. $2/\sigma$ individuals. Without this, there will be no chance of step two: recombination of a type 10 and a type 01 individual to produce a type 11. Finally, type 11 must become established, after which its ultimate fixation is essentially certain. Of course this may not happen the *first* time a new recombinant is produced. If type 11 becomes extinct and neither X_{10} nor X_{01} is one, then we can go back to step two.

In order to obtain a reasonable estimate for X_{10} when type 01 arises, we first examine the trajectory of X_{10} prior to the arrival of type 01. During this time, when X_{01} and X_{11} are both 0, we can write

$$X_{10}(s) = \frac{1}{2N} + M_{10}(s) + \int_0^s \sigma\gamma X_{10}(u)(1 - X_{10}(u))\, du,$$

where M_{10} is a mean-0 martingale with maximum jump size $1/(2N)$ and conditional quadratic variation

$$\langle M_{10}\rangle(s) = \frac{1+r}{2N}\int_0^s X_{10}(u)(1 - X_{10}(u))\, du,$$

i.e. $\langle M_{10}\rangle$ is the unique previsible process such that $M_{10}(s)^2 - M_{10}(0)^2 - \langle M_{10}\rangle(s)$ is a martingale (see e.g. §II.3.9 of Ikeda & Watanabe 1981). The additional factor of r in the martingale term $\langle M_{10}\rangle$, as compared to (1.1), is due to recombination between type 00 and type 10 individuals, which, because we have decoupled it from the resampling and selection, essentially increases the resampling rate. We drop the martingale term M_{10} and approximate the trajectory of X_{10} using a logistic growth curve, i.e.

$$X_{10}(s) \approx 1/(1 + (2N-1)\exp(-\sigma\gamma s)), \tag{2.2}$$

which solves $\frac{dX_{10}}{ds} = \sigma\gamma X_{10}(s)(1 - X_{10}(s))$ and $X_{10}(0) = 1/(2N)$. Notice that we are also dropping the coth term in (1.3), since its effect is non-negligible only when $X_{10} = \mathcal{O}(1/N)$, which lasts only a short time if N is large. As discussed in §1, we assume that the arrival time of the second mutation is uniformly distributed on the timecourse of the sweep (of the first mutation) and, since N is large, X_{10} spends most of the time near 0 or near 1. If we approximate the growth of X_{10} by the logistic growth curve (2.2), then it reaches $1/2$ at time $\frac{1}{\sigma\gamma}\log(2N-1) \approx \frac{1}{\sigma\gamma}\log(2N)$. Since we are most interested in the case when X_{10} is small, we choose the time of the introduction of the new mutation uniformly on $[0, \frac{1}{\sigma\gamma}\log(2N)]$. From now on this will be our time origin, so at $t = 0$ when the second mutation results in a single type 01 individual, $X_{10}(0) \approx (2N)^{-\zeta}$, where $\zeta \sim \mathrm{Unif}[0,1]$. In summary, at time $t = 0$ we take the state of the system to be

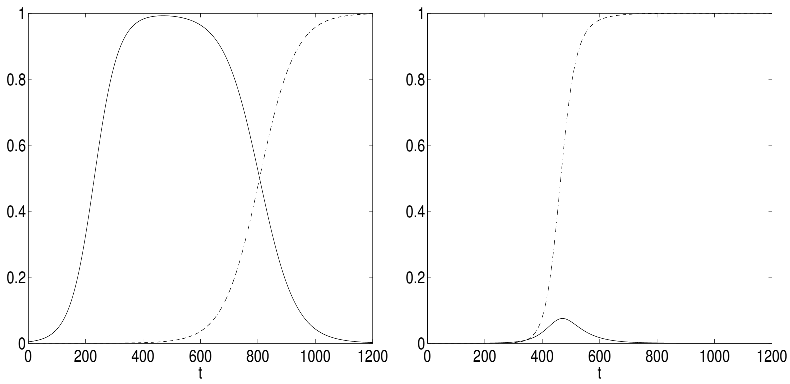
$$\begin{cases} X_{10}(0) = (2N)^{-\zeta} \text{ for some } \zeta \in [0,1] \\ X_{01}(0) = (2N)^{-1} \\ X_{11}(0) = 0 \end{cases}.$$

From this initial condition the first hurdle is for type 01 to become established, whose probability approaches $2\sigma/(1+\sigma)$ as $N \to \infty$ by approximating X_{01} using a branching process (see proof of Lemma 4.2(b) for more detail). Suppose that the first hurdle is indeed overcome and type 01 is established. From this time onwards, until type 11 becomes established, we approximate X_{01} and X_{10} deterministically. Until either is $\mathcal{O}(1)$, we have

$$X_{01}(t) \approx \frac{1}{2N}e^{\sigma t},\ X_{10}(t) \approx \frac{1}{(2N)^{\zeta}}e^{\sigma\gamma t}. \tag{2.3}$$

The above approximation fails once either X_{01} or X_{10} reaches $\mathcal{O}(1)$. We fix a small constant $c \ll 1$, then X_{01} reaches c approximately at time $\frac{1}{\sigma}(\log c + \log(2N))$ and X_{10} reaches c at approximately $\frac{1}{\sigma\gamma}(\log c + \zeta\log(2N))$. Our main result, Theorem 2.3, concerns the case when $\zeta < \gamma$. In this case, for sufficiently large N, X_{10} reaches c before X_{01}, and will further increase to almost 1 before X_{01} reaches c. At this time, which we denote T_1, the population consists almost entirely of types 10 or 01. Type 01, already established but still just a small proportion of the population, will then proceed to grow logistically, displacing type 10 individuals until X_{01} reaches $1-c$ at time T_2. The number of recombination events between X_{01} and X_{10} during $[T_1, T_2]$ is $\mathcal{O}(rN)$, which produces $\mathcal{O}(rN)$ (as $N \to \infty$ and then $c \to 0$) type 11 individuals, whereas the number of recombination events between X_{01} and X_{10} outside $[T_1, T_2]$ is $\mathcal{O}(rcN)$. Since c is small, we expect most recombination events to occur in $[T_1, T_2]$. Each type 11 individual has a probability of at least $2\sigma\gamma/(1+\sigma\gamma)$ of eventually becoming the common ancestor of all individuals in the population. So if we want to obtain a non-trivial limit (as $N \to \infty$) for the fixation probability of type 11, we should take $r = \mathcal{O}(1/N)$. When we use the term non-trivial here, we mean that as $N \to \infty$, (i) the fixation probability does not tend to 0, due to a lack of recombination events between type 01 and type 10 individuals, and (ii) nor does it tend to the establishment probability of type 01, due to infinitely many type 11 births, one of which is bound to sweep to fixation.

Consider for a moment the case $\zeta > \gamma$, where X_{01} reaches $\mathcal{O}(1)$ at time roughly $\frac{1}{\sigma}\log(2N)$, before X_{10} does, and X_{10} is $\mathcal{O}((2N)^{\gamma-\zeta})$ at this time. Furthermore, the



(a) $\zeta < \gamma$: $\zeta = 0.3, \gamma = 0.6$ (b) $\zeta > \gamma$: $\zeta = 0.7, \gamma = 0.6$

Figure 2: Approximate trajectories of X_{10} (solid line) and X_{01} (dashed line) when X_{11} is small: these curves are obtained assuming they undergo deterministic logistic growth with initial condition $X_{01}(0) = (2N)^{-1}$ and $X_{10}(0) = (2N)^{-\zeta}$. Parameter values: $\sigma = 0.02$, $(2N) = 10^8$. In Case 1 for $\zeta < \gamma$, X_{10} reaches almost 1 before being displaced by X_{01}, but in Case 2 where $\zeta > \gamma$, X_{10} never reaches $\mathcal{O}(1)$.

biggest X_{10} can get is $\mathcal{O}((2N)^{\gamma-\zeta})$ since X_{01} will very soon afterwards increase to almost 1, after which X_{10} will exponentially decrease (since type 10 is less fit than type 01). Hence we expect $\mathcal{O}(rN^{1+\gamma-\zeta})$ recombination events between type 01 and type 10, and the 'correct' scaling for r is $r = \mathcal{O}(N^{\zeta-\gamma-1})$ in this case. See Figure 2 for an explanation. We will examine the case of $\zeta > \gamma$ more thoroughly in §2.3.1.

We return to our main interest, the case $\zeta < \gamma$, which is the more likely scenario if γ is close to 1. We will take $r = \mathcal{O}(1/N)$ so that most of the recombination events between type 01 and type 10 individuals occur when type 01 is logistically displacing type 10, i.e. in the time interval $[T_1, T_2]$. We refine this by defining the constants in (4.1). The deterministic times t_{stoch}, t_{early}, t_{mid} and t_{late} roughly correspond to the lengths of the 'stochastic', 'early', 'middle', and 'late' phases of X_{01}, whose rôle is described in more detail in §4. Briefly, at the end of the stochastic phase, X_{01} is expected to reach either $\mathcal{O}((2N)^{a_0-1})$ or become extinct, where we will pick $a_0 = \zeta/(6\gamma)$ in §3. If X_{01} does not become extinct at the end of the stochastic phase, it becomes established and goes on to increase to level ϵ (at the beginning of §4, we will take $\epsilon = (2N)^{-(\gamma-\zeta)/144}$, which $\to 0$ as $N \to \infty$) during its early phase. During the middle phase of X_{01}, it increases from ϵ to $1-\epsilon$. After that, it enters its late phase, where t_{late} is chosen so that, with high probability, $X_{01} + X_{11}$ has fixed at 1 by the end of the late phase. In order to calculate the fixation probability of type 11, we only need to consider the evolution of X_{ij} while X_{11} is smaller than

$$\theta_{11} = \frac{\lceil \log(2N) \rceil}{2N},$$

since once X_{11} reaches θ_{11}, the fixation of type 11 is virtually certain (by (1.2), the probability that type 11 does not fix if it reaches θ_{11} is $< (2N)^{-2\sigma\gamma}$).

We consider the evolution of X_{01} after it reaches ϵ to be almost deterministic and approximate the subsequent evolution of X_{01} and X_{11} by Y_{01} and Y_{11}, defined below. Let

$$L(t; y_0, \beta) = \left[1 + \left(\frac{1}{y_0} - 1\right) e^{-\beta t}\right]^{-1} \tag{2.4}$$

be the solution to the logistic growth equation

$$L(t;y_0,\beta) = y_0 + \beta\int_0^t L(s;y_0,\beta)(1-L(s;y_0,\beta))\,ds.$$

We define Y_{01} to be $L(\cdot;\epsilon,\sigma(1-\gamma))$ during the middle phase of X_{01} and 1 during the late phase of X_{01}. Notice that $\sigma(1-\gamma)$ is the advantage of type 01 over type 10 and Y_{01} is deterministic. The time t_{mid} is exactly the length of time when Y_{01} is between ϵ and $1-\epsilon$. During the middle phase of X_{01}, we approximate the recombination events between type 01 and 10 individuals (which actually happen at rate $2rNX_{10}X_{01}$) as birth events of Y_{11} (which will happen at rate $2rNY_{01}(1-Y_{01})$). Recall that during the middle phase of X_{01} and as long as $X_{11}\le\theta_{11}$, $X_{01}+X_{10}$ is approximately 1. More precisely, for $t\in[0,t_{mid})$, we define

$$\begin{aligned}
Y_{01}(t) &= L(t;\epsilon,\sigma(1-\gamma)) \\
\mu^+(z,t) &= Nz[(1+\sigma(1+\gamma))(1-z)-(\sigma-r)Y_{01}(t)-(\sigma\gamma-r)(1-Y_{01}(t))] \\
&\quad +2rNY_{01}(t)(1-Y_{01}(t)) \\
\mu^-(z,t) &= Nz[(1-\sigma(1+\gamma)+2r)(1-z)+(\sigma-r)Y_{01}(t) \\
&\quad +(\sigma\gamma-r)(1-Y_{01}(t))],
\end{aligned} \tag{2.5}$$

and for $t\ge t_{mid}$, we define

$$\begin{aligned}
Y_{01}(t) &= 1 \\
\mu^+(z,t) &= N(1+\sigma\gamma+r)z(1-z) \\
\mu^-(z,t) &= N(1-\sigma\gamma+r)z(1-z).
\end{aligned} \tag{2.6}$$

We then take Y_{11} to be a birth and death process with birth and death rates $\mu^+(Y_{11},t)$ and $\mu^-(Y_{11},t)$, respectively, jump size $1/(2N)$, and initial condition $Y_{11}(0)=0$. It is absorbed on hitting θ_{11}. We couple Y_{11} to X_{11}, which jumps at rates r_{11}^+ and r_{11}^- defined in (2.1), in the standard manner, i.e. we use the same underlying Poisson process to construct both X_{11} and Y_{11}. More specifically, let $\Lambda_x^N(dt,d\phi)$ for $x\in\{01+,01-,10+,10-,11+,11-\}$ be a Poisson point process on $\mathbb{R}^+\times[0,1]$ with intensity measure $Ll\times l$, where l denotes Lebesgue measure and the constant L dominates the jump rates $r_{ij}^\pm$ in (2.1) and $\mu^\pm(z,t)$ in (2.5). For example, jumps of $\Lambda_{11+}^N(dt,d\phi)$ give possible times at which X_{11} increases by $1/(2N)$. Then X_{11} satisfies the following jump equation:

$$\begin{aligned}
X_{11}(t) &= X_{11}(0)+\frac{1}{2N}\int_0^t\int_0^1 1_{\phi\le r_{11}^+(s-)/L}\Lambda_{11+}^N(ds,d\phi) \\
&\quad -\frac{1}{2N}\int_0^t\int_0^1 1_{\phi\le r_{11}^-(s-)/L}\Lambda_{11-}^N(ds,d\phi),
\end{aligned}$$

and Y_{11} is constructed using the same Λ^N's:

$$\begin{aligned}
Y_{11}(t) &= X_{11}(0)+\frac{1}{2N}\int_0^t\int_0^1 1_{\phi\le\mu^+(Y_{11}(s-),s-)/L}\Lambda_{11+}^N(ds,d\phi) \\
&\quad -\frac{1}{2N}\int_0^t\int_0^1 1_{\phi\le\mu^-(Y_{11}(s-),s-)/L}\Lambda_{11-}^N(ds,d\phi).
\end{aligned} \tag{2.7}$$

With high probability, the trajectory of Y_{11} agrees with that of X_{11}. Its definition in terms of the jump rates $\mu^\pm$ takes into account both the birth of type 11 individuals due to recombination of type 01 and 10 individuals (second hurdle) and their subsequent establishment (third hurdle).

For $k \in \{0, 1/(2N), 2/(2N), \dots, \theta_{11}\}$, it is convenient to write $k_- = k - 1/(2N)$ and $k_+ = k + 1/(2N)$. We run Y_{11} until time $t_{mid} + t_{late}$. The probability that Y_{11} hits θ_{11} before then can be found by solving a system of ODE's. Let $p^{(11)}$ satisfy

$$\frac{d}{dt} p_k^{(11)}(t) = \begin{cases} \mu^-\left(1/(2N), t\right) p_{1/(2N)}^{(11)}(t) - \mu^+(0,t) p_0^{(11)}(t), & \text{if } k = 0 \\ \mu^+\left(\theta_{11,-}, t\right) p_{\theta_{11,-}}^{(11)}(t) - \mu^-(\theta_{11}, t) p_{\theta_{11}}^{(11)}(t), & \text{if } k = \theta_{11} \\ \mu^+\left(k_-, t\right) p_{k_-}^{(11)}(t) + \mu^-\left(k_+, t\right) p_{k_+}^{(11)}(t) & \\ \qquad -(\mu^+(k,t) + \mu^-(k,t)) p_k^{(11)}(t), & \text{otherwise} \end{cases} \tag{2.8}$$

with initial condition $p_k^{(11)}(0) = \mathbf{1}_{\{k=0\}}$, where $\theta_{11,-} = \theta_{11} - 1/(2N)$. Then

$$\mathbb{P}(Y_{11} \text{ hits } \theta_{11} \text{ before } t_{mid} + t_{late}) = p_{\theta_{11}}^{(11)}(t_{mid} + t_{late}). \tag{2.9}$$

The probability that X_{11} gets established, i.e. reaches θ_{11}, is then approximated by the probability that the birth and death process Y_{11} reaches θ_{11}. The latter can be found by solving the forward equation for the process Y_{11}, which can be found in (2.8).

Theorem 2.3. *If $\zeta < \gamma < 1$ and $r = \mathcal{O}(1/N)$, then there exist $\delta > 0$, whose value depends on σ, γ, and ζ, and a constant $C_{\gamma,\sigma}$, such that*

$$\left| \mathbb{P}\left(X_{11}(T_{fix}) = 1\right) - \frac{2\sigma}{1+\sigma} p_{\theta_{11}}^{(11)}(t_{mid} + t_{late}) \right| \le C_{\gamma,\sigma} N^{-\delta}$$

for sufficiently large N, where $p^{(11)}(t)$ solves the forward equation (2.8).

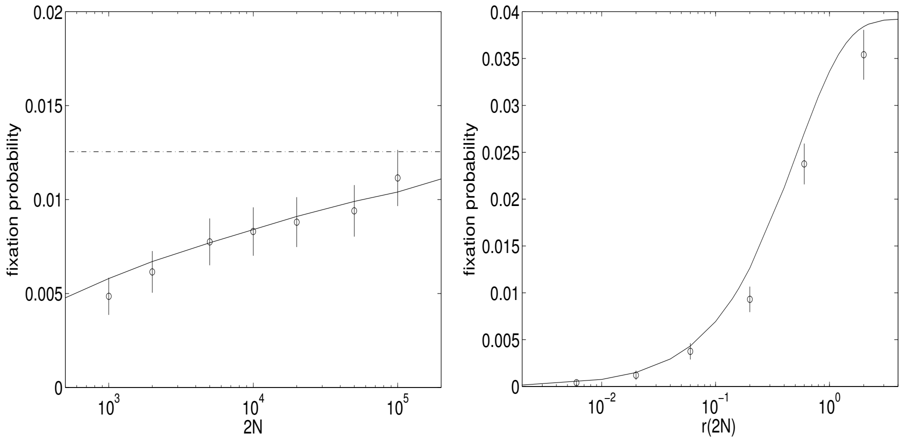
In the above, $\frac{2\sigma}{1+\sigma}$ corresponds to the probability of type 01 becoming established at the end of its stochastic phase, while $p_{\theta_{11}}^{(11)}(t_{mid} + t_{late})$ approximates the establishment probability of type 11 conditional on type 01 becoming established. Figure 3 compares fixation probabilities obtained from simulation, a non-rigorous calculation (which we briefly discuss in §2.3.2 below), and the large population limit of Theorem 2.3. In Figure 3(a) we hold rN constant in this simulation, and observe that the fixation probability of type 11 increases but does not change drastically as N becomes large. The reason for the drop in the fixation probability of type 11 when N is small may be because in this case, the early phase for X_{10} is very short and hence X_{10} increases to a level sufficient to reduce the establishment probability of type 01 before it actually gets established. In Figure 3(a), we use a population size of $2N = 50,000$ to approach the large population limit of Theorem 2.3. Apparently this population size still results in underestimates of the fixation probability of type 11 in the large population limit.

2.3 Other Cases

Our main result concerns the case $\zeta < \gamma$ when the second beneficial mutation occurs on a type 00 individual, and N is large. We briefly indicate how our results can be extended to some other cases.

2.3.1 The Case $\zeta > \gamma$

If γ is close to 1, this case is less likely than the case $\zeta < \gamma$, considered in Theorem 2.3. Nevertheless, we expect a similar result in this case, for which we provide an outline here. We take $\lambda \in (0, (\zeta - \gamma)/(2 - \gamma)]$ and $t'_{stoch} = \frac{1-\lambda}{\sigma} \log(2N)$, then at time t'_{stoch}, we expect X_{01} to be either 0 (with probability approximately $\frac{1-\sigma}{1+\sigma}$, as in the case $\zeta < \gamma$) or $\mathcal{O}((2N)^{-\lambda})$, and X_{10} to be roughly $(2N)^{(1-\lambda)\gamma - \zeta} \le (2N)^{-2\lambda}$. Since X_{10} and X_{11} can be



(a) (b)

Figure 3: Fixation probability of type 11: circles denote data points from simulations using 20,000 realisations with vertical bars denoting one standard deviation. (a) varying population size: the solid line denotes probabilities obtained using the non-rigorous calculation described in §2.3.2, and the dashed line denotes the large population limit of Theorem 2.3, with $r(2N) = 0.2$. (b) varying $r(2N)$: the solid line plots the large population limit of Theorem 2.3, and the simulation uses population size $2N = 50,000$. Other parameter values: $\sigma = 0.02$, $\zeta = 0.3$ and $\gamma = 0.6$.

expected to be quite small before t'_{stoch}, they exert little influence on the trajectory of X_{01}, which jumps by $\pm 1/(2N)$ at roughly the following rates:

$$r_{01}^{+} \approx N(1+\sigma+r)X_{01},\ r_{01}^{-} \approx N(1-\sigma+r)X_{01}.$$

Hence before t'_{stoch}, $2NX_{01}$ resembles a continuous-time branching process Z with offspring generating function $u(s) = \frac{1}{2}(1+\sigma+r)s^2 + \frac{1}{2}(1-\sigma+r) - (1+r)s$. Using Theorem III.8.3 of Athreya & Ney (1972), we can calculate $E[e^{-uW}]$ for $W = \lim_{t\to\infty} e^{-\sigma t}Z(t)$ and conclude that W is distributed according to $\frac{1-\sigma+r}{1+\sigma+r}\delta_0(y) + \exp(-\frac{2\sigma}{1+\sigma+r}y)\ dy$ for $y \geq 0$. Hence the conditional distribution function of $X_{01}(t'_{stoch})|X_{01}(t'_{stoch}) > 0$ resembles $Exp(\frac{1+\sigma+r}{2\sigma}(2N)^{-\lambda})$, an exponential distribution with mean $\frac{1+\sigma+r}{2\sigma}(2N)^{-\lambda}$, as $N \to \infty$.

From time t'_{stoch} onwards, until either X_{01} gets very close to 0 or X_{10} becomes much smaller than $\mathcal{O}((2N)^{(1-\lambda)\gamma-\zeta})$, we can assume that the paths of X_{10} and X_{01} are well approximated by those of Y_{10} and Y_{01}, respectively, where

$$\begin{aligned} dY_{01} &= Y_{01}[(1+\sigma)(1-Y_{01}) - \sigma\gamma Y_{10}]\ dt \\ dY_{10} &= Y_{10}[(1+\sigma\gamma)(1-Y_{10}) - \sigma Y_{01}]\ dt \end{aligned}$$

with the initial condition $Y_{01}(0)$ (corresponding to $X_{01}(t'_{stoch})$) drawn from an $Exp(\frac{1+\sigma+r}{2\sigma}(2N)^{-\lambda})$ distribution and $Y_{10}(0) = (2N)^{(1-\lambda)\gamma-\zeta}$. As in the case $\zeta < \gamma$, we can then approximate X_{11} by a birth and death process Y_{11} with rates the same as $r_{11}^{\pm}$ from (2.1) but with X_{01} replaced by Y_{01} and X_{10} replaced by Y_{10}. Let $p_{\theta_{11}}^{11}(t;y)$ be the probability that Y_{11} reaches θ_{11} by time t, then $p_{\theta_{11}}^{11}(t;y)$ can be found by solving the forward equation for Y_{11}, i.e. (2.8) but with the following μ^+ and μ^-:

$$\begin{aligned} \mu^{+}(z,t) &= Nz[(1+\sigma(1+\gamma))(1-z) - (\sigma-r)Y_{01}(t) - (\sigma\gamma-r)Y_{10}(t)] \\ &\quad +2rNY_{01}(t)Y_{10}(t) \\ \mu^{-}(z,t) &= Nz[(1-\sigma(1+\gamma)+2r)(1-z) + (\sigma-r)Y_{01}(t) + (\sigma\gamma-r)Y_{10}(t))]. \end{aligned}$$

Finally, we average over the distribution of $Y_{01}(0)$, and

$$\int_0^\infty \lim_{t\to\infty} p^{11}_{\theta_{11}}(t;y)e^{-\frac{2\sigma}{1+\sigma+r}y}\,dy$$

can be expected to be within $N^{-\delta}$ of the fixation probability of type 11 for some $\delta > 0$. The proof of such a result is more tedious than that of Theorem 2.3 but makes use of similar ideas.

2.3.2 Brief Comment on Moderate N

For moderate population sizes, the observation that X_{10} increases to close to 1 before X_{01} reaches $\mathcal{O}(1)$ breaks down. We can, however, compute the distribution function $f_{T_{01;\theta_{01}}}$ of the random time $T_{01;\theta_{01}}$ when X_{01} hits a certain level θ_{01}, assuming that X_{10} grow logistically before $T_{01;\theta_{01}}$. From $T_{01;\theta_{01}}$ onwards and before X_{11} hits θ_{11}, X_{01} grows roughly deterministically, displacing both type 10 and type 00, so we can approximate X_{11} by Y_{11}, a birth and death process with time-varying jump rates in the form of $r^{\pm}_{11}$ in (2.1), but with X_{01}, X_{10} and X_{00} replaced by their deterministic approximations. Assuming $T_{01;\theta_{01}} = t$, we can numerically solve the forward equation for Y_{11}, which is directly analogous to (2.8), to find the probability that Y_{11} eventually hits θ_{11}, which we denote by $p^{(11)}_{est}(t)$. The dependence of $p^{(11)}_{est}$ on t comes through the initial condition X_{10} for the ODE system, which depends on $T_{01;\theta_{01}}$. The fixation probability of type 11 is then approximately $\int p^{(11)}_{est}(t) f_{T_{01;\theta_{01}}}(t)\,dt$. This is the algorithm we use to produce the solid line in Figure 3(a) and is given in its full detail in Yu & Etheridge (2010).

3 Proof of the Main Theorem

We first define some of the functions and events needed for the proof, then give some intuition, before we proceed with the proof of Theorem 2.3. We define

$$\begin{aligned} T_{Z;x} &= \inf\{t \geq 0 : Z \geq x\} \\ T_{ij;x} &= \inf\{t \geq 0 : X_{ij} \geq x\} \end{aligned} \tag{3.1}$$

for any $ij \in \{00, 10, 01, 11\}$. We also define the stopping time

$$T_\infty = T_{01;\epsilon} + t_{mid} + t_{late}.$$

Recall that the deterministic times t_{stoch}, t_{early}, t_{mid} and t_{late} roughly correspond to the lengths of the 'stochastic', 'early', 'middle', and 'late' phases of X_{01}, whose rôle is described in more detail in §4. We define

$$\begin{aligned} a_0 = \frac{\zeta}{6},\ t_{stoch} = \frac{a_0}{\sigma}\log(2N),\ t_{early} = \frac{1.01\log(2N)}{\sigma(1-\gamma)-r}. \\ t_{mid} = \frac{1}{\sigma(1-\gamma)}\log\frac{1-\epsilon}{\epsilon},\ t_{late} = \frac{1.02}{\sigma\gamma}\log(2N), \end{aligned} \tag{3.2}$$

At the end of the stochastic phase, X_{01} is expected to reach either $\mathcal{O}((2N)^{a_0-1})$ or become extinct. Also recall the definition of Y_{01} and Y_{11} in (2.5) and (2.6). We define

$$\begin{aligned} Z_{01}(t) &= \begin{cases} 0, & \text{if } t < T_{01;\epsilon} \\ Y_{01}(t - T_{01;\epsilon}), & \text{if } t \geq T_{01;\epsilon} \end{cases} \\ Z_{11}(t) &= \begin{cases} 0, & \text{if } t < T_{01;\epsilon} \\ Y_{11}(t - T_{01;\epsilon}), & \text{if } t \geq T_{01;\epsilon} \end{cases} \end{aligned} \tag{3.3}$$

With the convention of (3.1),

$$T_{Z_{01};1-\epsilon} = T_{01;\epsilon} + t_{mid},$$

and we observe that $Z_{01}(t) = 1$ for $t \geq T_{Z_{01};1-\epsilon}$. Notice that Z_{01} is deterministic other than the time it becomes nonzero (i.e. at $T_{01;\epsilon}$) and the jumps of Z_{11} are coupled to those of X_{11}, since Z_{11} is simply a time shifted version of Y_{11}, defined after (2.6).

We define the following events:

$$\begin{aligned}
E_0 &= \{T_{Z_{01};1-\epsilon} \leq T_{11;\theta_{11}}\} \\
E_1 &= \{X_{01}(t_{stoch}) > 0\} \\
G_1 &= \{T_{01;\epsilon} \leq T_{11;1/(2N)} \wedge (t_{stoch} + t_{early})\} \cap \{X_{10}(T_{01;\epsilon}) \geq 1 - \epsilon - \epsilon^4\} \\
G_2 &= \{|X_{01}(t) - Z_{01}(t)| \leq \epsilon^{1/3} \text{ and } X_{00}(t) \leq \epsilon^3 \text{ for all} \\
&\qquad t \in [T_{01;\epsilon}, T_{Z_{01};1-\epsilon} \wedge T_{11;\theta_{11}}]\} \\
G_3 &= \{X_{11}(t) + X_{01}(t) > 1 - \epsilon^{1/2} \text{ for all } t \geq T_{Z_{01};1-\epsilon}\} \\
G_4 &= \{X_{11}(T_\infty) + X_{01}(T_\infty) = 1\} \\
G_5 &= \{X_{11}(t) = Z_{11}(t) \text{ for all } t \in [T_{01;\epsilon}, T_\infty \wedge T_{11;\theta_{11}}]\} \\
G_6 &= \{T_{11;\theta_{11}} \leq T_\infty \text{ or } X_{11}(T_\infty) = Z_{11}(T_\infty) = 0\}.
\end{aligned} \tag{3.4}$$

On event E_0, type 11 is not established before $T_{Z_{01};1-\epsilon}$, which roughly corresponds to the event that type 11 never gets established since there will be very few recombination events after $T_{Z_{01};1-\epsilon}$. On event E_1, the second mutation survives the initial stochastic phase. Events E_0 and E_1 have probabilities that do not approach 0 or 1 as $N \to \infty$. Conditional on event E_1, events G_1 to G_6 occur with high probability, which enable us to approximate X_{01} and X_{11} by Z_{01} and Z_{11}, respectively. We outline the intuition behind these definitions in the following three paragraphs and refer to readers to Figure 4 and Table 1 for an illustration of different phases of X_{10} and X_{01}. Recall that t_{stoch} is the length of the initial 'stochastic' phase for X_{01}. At t_{stoch}, with high probability X_{01} either is $\mathcal{O}((2N)^{a_0-1})$ or has hit 0 (event E_1^c). In the latter case, there is no need to approximate X_{01} any further. On the other hand, if E_1 occurs, then type 01 is very likely to be established by t_{stoch} and, with high probability, have overcome the first hurdle to the fixation of type 11 outlined at the beginning of §2.2. Thus it grows almost deterministically to reach level ϵ (slightly smaller than $\mathcal{O}(1)$) at time $T_{01;\epsilon}$ (first half of G_1). Note that t_{early} is picked so that X_{01}, which has advantage $\sigma(1-\gamma)$ over type 10, can grow from roughly $\mathcal{O}(1/N)$ to at least $\mathcal{O}(1)$ in time t_{early}. Furthermore, as discussed in §1, since $\zeta < \gamma$, with high probability X_{10} will grow to around ϵ^{24} at time $t_{10;\epsilon^{24}}$ and further increase to close to 1 at time $t'_{10;1-\epsilon^{24}}$. With high probability, $t'_{10;1-\epsilon^{24}} < T_{01;\epsilon}$ so that the second half of G_1 is also likely (please refer to (4.2) for exact definition of $t_{10;\epsilon^{24}}$ and $t'_{10;1-\epsilon^{24}}$). Hence conditional on E_1, event G_1 is very likely.

After $T_{01;\epsilon}$, X_{01} enters its middle phase and can be well approximated by the deterministic process Z_{01} (first half of G_2), which grows logistically at rate $\sigma(1-\gamma)$ until $T_{Z_{01};1-\epsilon}$. This approximation requires X_{11} to remain small (event E_0). Furthermore, since $X_{10}(T_{01;\epsilon}) \approx 1$, from $T_{01;\epsilon}$ onwards, X_{00} is likely to remain small since type 00 is the least fit type (second half of G_2). During $[T_{01;\epsilon}, T_{Z_{01};1-\epsilon}]$, the definition of Z_{11} takes into account recombination events between type 10 and 01 individuals that produce type 11 individuals at a rate of $2rNX_{10}X_{01}$, which is approximated by $2rNZ_{01}(1-Z_{01})$ in the definition of Z_{11}. As discussed at the beginning of §2.2, the second hurdle to the fixation of type 11 is the production of a type 11 individual via recombination between

X_{10}	start time	start position	approximation
early	0	$(2N)^{-\zeta}$	$L(\cdot;(2N)^{-\zeta},\sigma\gamma)$
middle	$t_{10;\epsilon^{24}}$	around ϵ^{24}	$L(\cdot - t_{10;\epsilon^{24}}; X_{10}(t_{10;\epsilon^{24}}),\sigma\gamma)$
late	$t'_{10;1-\epsilon^{24}}$	around $1-\epsilon^{24}$	$X_{01}+X_{10}$ stays above $1-\epsilon^4$
X_{01}	start time	start position	approximation
stochastic	0	$(2N)^{-1}$	branching process defined in Lemma 4.2(b)
early	t_{stoch}	around $(2N)^{a_0-1}$	$Z_{01}=0$, upper and lower bounds in Lemma 4.2(c,d)
middle	$T_{01;\epsilon}$	ϵ	$Z_{01}(T_{01;\epsilon}+\cdot)=L(\cdot;\epsilon,\sigma(1-\gamma))$
late	$T_{Z_{01};1-\epsilon}$	around $1-\epsilon$	$Z_{01}=1$

Table 1: Different phases of X_{10} and X_{01}: X_{01} and X_{11} are approximated by Z_{01} and Z_{11}, respectively, during the early, middle and late phases of X_{01}. Z_{01} and Z_{11} are time-shifted versions of Y_{01} and Y_{11}, defined in (2.5-2.7)

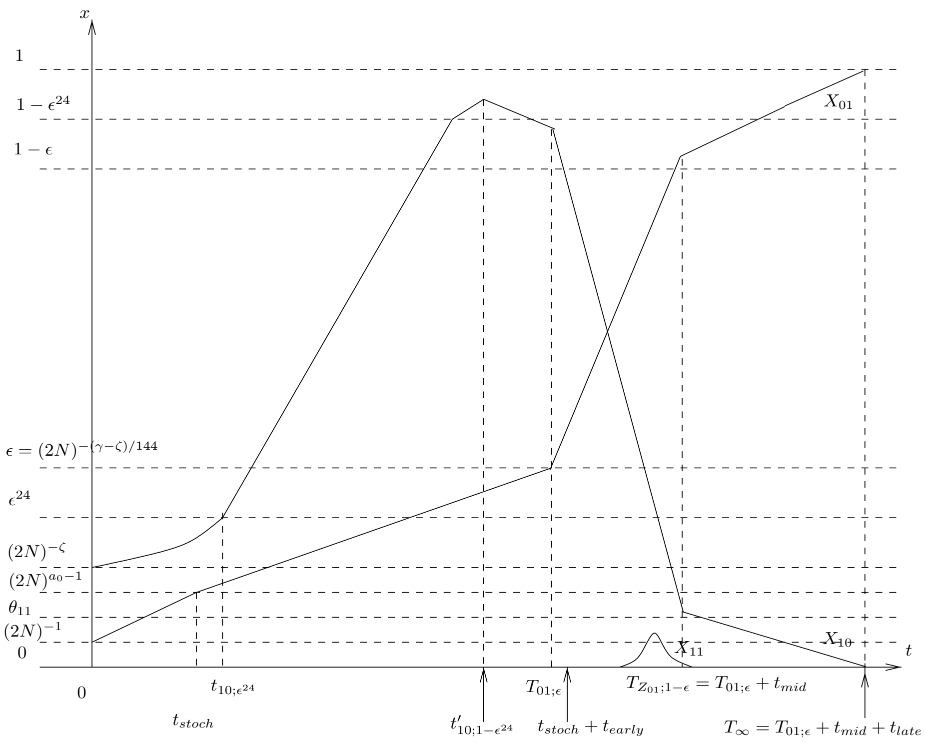


Figure 4: Illustration of trajectories of X_{ij}: in this particular example, X_{11} never establishes and X_{01} sweeps to fixation. Neither the t nor the x axis is to scale.

type 01 and 10 individuals. Notice that we can approximate X_{10} by $1 - Z_{01}$ since we assume X_{00} is small under G_2 and X_{11} is small throughout. As a result, the jump process Z_{11} closely approximates X_{11}, conditional on $G_2 \cap G_1 \cap E_1$ (event G_5). This approximation remains valid until either when type 11 becomes established or the end of the late phase at T_∞.

After $T_{Z_{01};1-\epsilon}$, X_{01} enters its late phase and we just need to make sure that with high probability we obtain a definite answer on whether X_{11} has reached θ_{11} or 0 by time T_∞. For this, we need $X_{11} + X_{01}$ to remain close to 1 (event G_3) and hit 1 at time T_∞ (event G_4). If Z_{11} has not hit θ_{11} by the beginning of the late phase of X_{01} at time $T_{Z_{01};1-\epsilon}$ (event E_0), then we continue to keep track of Z_{11} until the end of the late phase at T_∞, when it most likely has already hit either θ_{11} or 0 (event G_6). We ignore any more recombination events between type 01 and 10 and Z_{11} is a time-changed branching process during the late phase of X_{01}. If X_{11} hits 0 by T_∞, then we regard type 11 as having failed to establish and since X_{01} is most likely to be 1 (event G_4) at T_∞, the earlier mutation has gone extinct. On the other hand, if X_{11} hits θ_{11} by T_∞, then we regard type 11 as having overcome the third hurdle to its fixation and become established. Hence it will, with high probability, eventually sweep to fixation (Lemma 3.2).

Proposition 3.1 below estimates the probabilities of the 'good' events G_1 through G_6. It is essential for the proof of Theorem 2.3, and will be proved in §4.

Proposition 3.1. *If $\zeta < \gamma < 1$ and $r = \mathcal{O}(1/N)$, then there exist positive constants δ whose exact value depends on σ, γ and ζ, and a constant $C_{\gamma,\sigma}$, such that for sufficiently large N,*

$$\begin{aligned}
&(a) \quad \left|\mathbb{P}(E_1^c) - \frac{1-\sigma+r}{1+\sigma+r}\right| \le C_{\gamma,\sigma} N^{-\delta} \\
&(b) \quad \mathbb{P}(G_1^c \cap E_1) \le C_{\gamma,\sigma} N^{-\delta} \\
&(c) \quad \mathbb{P}(G_2^c \cap G_1 \cap E_1) \le C_{\gamma,\sigma} N^{-\delta} \\
&(d) \quad \mathbb{P}(G_3^c \cap E_0 \cap G_2 \cap G_1 \cap E_1) \le C_{\gamma,\sigma} N^{-\delta} \\
&(e) \quad \mathbb{P}\left(G_4^c \cap E_0 \cap G_2 \cap G_1 \cap E_1\right) \le C_{\gamma,\sigma} N^{-\delta}.
\end{aligned}$$

Consequently, we have (f) $\mathbb{P}\left(G_4^c \cap E_0 \cap G_1 \cap E_1\right) \le 2C_{\gamma,\sigma} N^{-\delta}$. Furthermore,

$$\begin{aligned}
&(g) \quad \mathbb{P}(G_5^c \cap G_1 \cap E_1) \le C_{\gamma,\sigma} N^{-\delta} \\
&(h) \quad \mathbb{P}(G_6^c \cap G_5 \cap G_1 \cap E_1) \le C_{\gamma,\sigma} N^{-\delta}.
\end{aligned}$$

To establish part (f) of above, we observe that

$$\begin{aligned}
&\mathbb{P}\left(G_4^c \cap E_0 \cap G_1 \cap E_1\right) \\
&\quad = \mathbb{P}\left(G_4^c \cap E_0 \cap G_2 \cap G_1 \cap E_1\right) + \mathbb{P}\left(G_4^c \cap E_0 \cap G_2^c \cap G_1 \cap E_1\right) \\
&\quad \le \mathbb{P}\left(G_4^c \cap E_0 \cap G_2 \cap G_1 \cap E_1\right) + \mathbb{P}\left(G_2^c \cap G_1 \cap E_1\right),
\end{aligned}$$

which can be estimated using (e) and (c), respectively. We will also need to show that the probability of type 11 becoming fixed is approximately the probability that it reaches θ_{11}.

Lemma 3.2. $|\mathbb{P}(X_{11}(T_{fix}) = 1) - \mathbb{P}(T_{11;\theta_{11}} < \infty)| \le (2N)^{\log \frac{1-\sigma\gamma+2r}{1+\sigma\gamma}}$.

Proof. On $\{T_{11;\theta_{11}} < \infty\}$, X_{11} dominates $\check{X}_{11}$, a birth and death process with initial condition $\check{X}_{11}(T_{11;\theta_{11}}) = \theta_{11} = \lceil \log(2N) \rceil/(2N)$, jump size $1/(2N)$, and the following jump rates

$$\check{r}_{11}^+ = N(1+\sigma\gamma)\check{X}_{11}(1-\check{X}_{11}),\ \check{r}_{11}^- = N(1-\sigma\gamma+2r)\check{X}_{11}(1-\check{X}_{11}).$$

Using standard Markov chain techniques (see e.g. Example 3.9.6 of Grimmett & Stirzaker 1992), we may conclude that the probability that $\check{X}_{11}$ first hits 0 instead of 1 is

$$\frac{(\frac{1-\sigma\gamma+2r}{1+\sigma\gamma})^{\log 2N} - (\frac{1-\sigma\gamma+2r}{1+\sigma\gamma})^{2N}}{1-(\frac{1-\sigma\gamma+2r}{1+\sigma\gamma})^{2N}} \leq \left(\frac{1-\sigma\gamma+2r}{1+\sigma\gamma}\right)^{\log 2N} = (2N)^{\log \frac{1-\sigma\gamma+2r}{1+\sigma\gamma}}$$

since $(x^{\log 2N} - x^{2N})/(1-x^{2N}) < x^{\log 2N}$ for $x \in (0,1)$, hence

$$\mathbb{P}(\{T_{\check{X}_{11};1} > T_{\check{X}_{11};0}, T_{11;\theta_{11}} < \infty\}) \leq (2N)^{\log \frac{1-\sigma\gamma+2r}{1+\sigma\gamma}},$$

which implies $\mathbb{P}(\{X_{11}(T_{fix}) \neq 1,\ T_{11;\theta_{11}} < \infty\}) \leq (2N)^{\log \frac{1-\sigma\gamma+2r}{1+\sigma\gamma}}$. Since for any two events A and B

$$|\mathbb{P}(A) - \mathbb{P}(B)| \leq \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B)$$

and $\{X_{11}(T_{fix}) = 1,\ T_{11;\theta_{11}} = \infty\}$ is a set with probability 0, we have the desired result. □

Proof of Theorem 2.3. **Step 1: E_1^c can be ignored.** In this step, we show that there is almost no hope of fixation for type 11 on E_1^c, where X_{01} has hit 0 by the end of the stochastic phase. Let

$$G_7 = \{X_{11}(t) = 0 \text{ for all } t \leq t_{stoch}\}.$$

Comparing with (2.1), we see that the jump process $\hat{X}_{01}$ with initial condition $\hat{X}_{01}(0) = 1/(2N)$, jump size $1/(2N)$, and the following jump rates

$$\hat{r}_{01}^{+} = N(1+\sigma)\hat{X}_{01} + 3rN,\ \hat{r}_{01}^{-} = N(1-\sigma)\hat{X}_{01}$$

dominates X_{01} for all time. Then

$$d\hat{X}_{01} = dM + (\sigma\hat{X}_{01} + \frac{3}{2}r)\,dt$$

where M is a martingale with maximum jump size $1/(2N)$ and conditional quadratic variation $\langle M\rangle$ satisfying $d\langle M\rangle = \frac{1}{2N}(2\hat{X}_{01} + 3r)\,dt$. Hence

$$E[\hat{X}_{01}(t)] = \left(\frac{1}{2N} + \frac{3r}{2\sigma}\right)e^{\sigma t} - \frac{3r}{2\sigma} \leq \left(\frac{1}{2N} + \frac{3r}{2\sigma}\right)e^{\sigma t},$$

We recall Burkholder's inequality in the following form:

$$E\left[\sup_{s\leq t}|M(s)|^p\right] \leq C_p\mathbb{E}\left[\langle M\rangle(t)^{p/2} + \sup_{s\leq t}|M(s) - M(s-)|^p\right]$$

for $p \geq 1$, which may be derived from its discrete time version, Theorem 21.1 of Burkholder (1973). We use this and Jensen's inequality to obtain

$$\begin{aligned} E\left[\sup_{s\leq t_{stoch}}|M(s)|\right] &\leq E\left[\sup_{s\leq t_{stoch}}|M(s)|^2\right]^{1/2} \\ &\leq \frac{C}{N}\left(1 + N\int_0^{t_{stoch}} E[\hat{X}_{01}(s) + \frac{3r}{2}]\,ds\right)^{1/2} \\ &\leq \frac{C}{N} + \frac{C_\sigma}{\sqrt{N}}\left(rt_{stoch} + (N^{-1}+r)e^{\sigma t_{stoch}}\right)^{1/2} \\ &\leq C_\sigma N^{(a_0/2)-1}. \end{aligned} \tag{3.5}$$

where we use the assumption $r = \mathcal{O}(1/N) < C/N$ and the definition of the deterministic time t_{stoch} from (3.2). Therefore

$$\begin{aligned} E\left[\sup_{s\le t_{stoch}} \hat{X}_{01}(s)\right] &\le E\left[\sup_{s\le t_{stoch}} |M(s)|\right] + \frac{3r}{2}t_{stoch} + \sigma\int_0^{t_{stoch}} E[\hat{X}_{01}(s)]\,ds \\ &\le C_\sigma N^{a_0-1}, \end{aligned}$$

because $\int_0^{t_{stoch}} E[\hat{X}_{01}(s)]\,ds$ dominates. Since $\hat{X}_{01}$ dominates X_{01}, we have

$$\mathbb{P}\left(\sup_{s\le t_{stoch}} X_{01}(s) \ge (2N)^{2a_0-1}\right) \le C_\sigma N^{-a_0}.$$

On $\{\sup_{s\le t_{stoch}} X_{01}(s) < (2N)^{2a_0-1}\}$, the number of recombination events between type 01 and 10 during $[0, t_{stoch}]$ is at most $\mathrm{Poisson}(2r(2N)^{2a_0-1}t_{stoch})$, hence

$$\begin{aligned} \mathbb{P}(G_7^c \cap E_1^c) \le \mathbb{P}(G_7^c) &\le \mathbb{P}\left(G_7^c \cap \left\{\sup_{s\le t_{stoch}} X_{01}(s) < (2N)^{2a_0-1}\right\}\right) \\ &\quad +\mathbb{P}\left(\sup_{s\le t_{stoch}} X_{01}(s) \ge (2N)^{2a_0-1}\right) \\ &\le C_\sigma(N^{-a_0} + N^{(2a_0-1)/2}) \end{aligned}$$

for sufficiently large N. On $G_7 \cap E_1^c$, type 01 has gone extinct by time t_{stoch}, before a single individual of type 11 has been born, hence type 11 will not get established, let alone fix. Therefore

$$\mathbb{P}\left(\{T_{11;\theta_{11}} < \infty\} \cap E_1^c\right) \le \mathbb{P}(G_7^c \cap E_1^c) \le C_{r,\sigma}(N^{-a_0} + N^{(2a_0-1)/2}). \tag{3.6}$$

Step 2: On E_1, everything is decided by T_∞. Now we concentrate on E_1 where type 01 has most likely established itself at time t_{stoch}. The event whose probability does not approach 0 when $N \to \infty$ is $G_6 \cap G_5 \cap G_1 \cap E_1$. We split G_6 into two disjoint events:

$$\begin{aligned} G_{61} &= \{T_{11;\theta_{11}} \le T_\infty\} \\ G_{62} &= \{T_{11;\theta_{11}} > T_\infty, X_{11}(T_\infty) = Z_{11}(T_\infty) = 0\}. \end{aligned}$$

On $E_1 \cap G_1$, for exactly one of the two events $\{T_{11;\theta_{11}} < \infty\}$ and $\{T_{Z_{11};\theta_{11}} \le T_\infty\}$ to occur (i.e. either the former occurs but the latter does not, or the latter occurs and the former does not), one of the following three scenarios must occur:

1. X_{11} and Z_{11} disagree before $T_\infty \wedge T_{11;\theta_{11}}$, i.e. G_5^c;
2. X_{11} and Z_{11} agree up to T_∞, but do not hit $\{0, \theta_{11}\}$ before T_∞, i.e. $G_6^c \cap G_5$;
3. X_{11} and Z_{11} agree up to T_∞, X_{11} does not hit θ_{11} before T_∞ and $X_{11}(T_\infty) = 0$, but $X_{01}(T_\infty) < 1$ thus allowing the possibility of type 11 being born due to recombination between type 01 and 10 individuals after T_∞, i.e $G_{62} \cap G_5 \cap G_4^c$.

By Prop 3.1(b), Prop 3.1(g-h), and Prop 3.1(f), respectively, we can pick $\delta > 0$ (whose value may change line from line) such that

$$\begin{aligned} \mathbb{P}(G_1^c \cap E_1) &\le C_{\gamma,\sigma}N^{-\delta} \\ \mathbb{P}((G_6^c \cup G_5^c) \cap G_1 \cap E_1) &\le C_{\gamma,\sigma}N^{-\delta} \\ \mathbb{P}(G_{62} \cap G_5 \cap G_4^c \cap G_1 \cap E_1) &\le C_{\gamma,\sigma}N^{-\delta}, \end{aligned} \tag{3.7}$$

where the last estimate above follows from the fact $G_{62} \subset E_0$, since $T_\infty \ge T_{Z_{01};1-\epsilon} = T_{01;\epsilon} + t_{mid}$.

For any two events A and B, $|\mathbb{P}(A \cap E) - \mathbb{P}(B \cap E)| \leq \mathbb{P}(A \cap B^c \cap E) + \mathbb{P}(A^c \cap B \cap E)$, therefore

$$\begin{aligned} &|\mathbb{P}\left(\{T_{11;\theta_{11}} < \infty\} \cap E_1\right) - \mathbb{P}\left(\{T_{Z_{11};\theta_{11}} \leq T_\infty\} \cap E_1\right)| \\ &\quad \leq \; 2\mathbb{P}(G_1^c \cap E_1) + \mathbb{P}((G_6^c \cup G_5^c \cup (G_{62} \cap G_5 \cap G_4^c)) \cap G_1 \cap E_1) \\ &\quad \leq \; C_{\gamma,\sigma} N^{-\delta} \end{aligned}$$

by (3.7). From (3.6), we have

$$\begin{aligned} &|\mathbb{P}\left(T_{11;\theta_{11}} < \infty\right) - \mathbb{P}\left(\{T_{11;\theta_{11}} < \infty\} \cap E_1\right)| \\ &\quad = \; \mathbb{P}\left(\{T_{11;\theta_{11}} < \infty\} \cap E_1^c\right) \leq C_\sigma N^{-a_0} + N^{(2a_0-1)/2}. \end{aligned}$$

But by Proposition 3.1(a),

$$\left| \mathbb{P}(E_1) - \frac{2\sigma}{1+\sigma} \right| \leq C_{\gamma,\sigma} N^{-\delta}.$$

We combine the three inequalities above to conclude

$$\begin{aligned} &\left| \mathbb{P}\left(T_{11;\theta_{11}} < \infty\right) - \frac{2\sigma}{1+\sigma} \mathbb{P}\left(T_{Z_{11};\theta_{11}} \leq T_\infty | E_1\right) \right| \\ &\quad \leq \; |\mathbb{P}\left(T_{11;\theta_{11}} < \infty\right) - \mathbb{P}\left(T_{Z_{11};\theta_{11}} \leq T_\infty | E_1\right) \mathbb{P}(E_1)| + C_{\gamma,\sigma} N^{-\delta} \\ &\quad = \; |\mathbb{P}\left(T_{11;\theta_{11}} < \infty\right) - \mathbb{P}\left(\{T_{Z_{11};\theta_{11}} \leq T_\infty\} \cap E_1\right)| + C_{\gamma,\sigma} N^{-\delta} \\ &\quad \leq \; C_{\gamma,\sigma} N^{-\delta} \end{aligned}$$

for some $\delta > 0$, and then use Lemma 3.2, as well as (2.9) and (3.3) to obtain the desired conclusion. □

4 Proof of Proposition 3.1

We divide the evolution of X_{01} and X_{10} roughly into 4 phases, 'stochastic', 'early', 'middle', and 'late', and use Lemmas 5.1, 5.2, and 5.3 for each of the last 3 phases, respectively. The phases of X_{01} and X_{10} are not concurrent, e.g. under the assumption $\zeta < \gamma < 1$, X_{10} will have entered its late phase before X_{01} finishes the early phase. See Figure 4 for an illustration of the different phases of X_{01} and X_{10}. Lemma 4.1 deals with the early, middle, and late phases of X_{10}. Because X_{10} starts at $U = (2N)^{-\zeta} \gg 1/(2N)$ at $t = 0$, it has no stochastic phase. Its early phase is between $t = 0$ and the time when X_{10} reaches ϵ^{24}. Its middle phase is between ϵ^{24} and $1 - \epsilon^{24}$, after which it enters the late phase.

Lemma 4.2 deals with the stochastic and early phases of X_{01}. Since $X_{01}(0) = 1/(2N)$, whether it establishes itself is genuinely stochastic (i.e. its probability tends to a positive constant strictly less than 1 as $N \to \infty$). The stochastic phase lasts for time t_{stoch}, when, with high probability, either type 01 has established or it has gone extinct. If X_{01} reaches $\mathcal{O}((2N)^{a_0-1})$ by time t_{stoch}, it enters the early phase. Part (b) of Lemma 4.2 says that if $\kappa = \zeta/\gamma < 1$ then it does not reach $(2N)^{-(1-\kappa)/2}$ until X_{10} has entered its late phase, while part (c) says that it does reach $\epsilon = (2N)^{-(\gamma-\zeta)/144} > (2N)^{-(1-\kappa)/2}$ at some finite time.

With Lemmas 4.1 and 4.2 in hand, we move on to the proof of Proposition 3.1. The proof of parts (a-b) of this proposition reconciles stopping times used in Lemmas 4.1 and 4.2 and establishes that X_{01} is likely to reach ϵ if it becomes established, and when that happens, X_{10} is already close to 1. It then goes on to establish that every 'good' event defined in (3.4) is likely to occur.

Recall the definition of the logistic growth curve $L(t; y_0, \beta)$ from (2.4). Throughout the rest of this section, we use $L(t; (2N)^{-\zeta}, \sigma\gamma)$ to approximate the trajectory of X_{10}

during its early phase and let

$$t_{10;x} = \inf\{t \geq 0 : L(t; (2N)^{-\zeta}, \sigma\gamma) \geq x\}$$

(with the stipulation that $t_{10;x} = 0$ if $(2N)^{-\zeta} \geq x$), e.g. $t_{10;\epsilon^{24}}$ is when this approximation hits ϵ^{24}. Furthermore, we use $t_{10,x,y}$ to denote the time this approximation spends between x and y. Thus

$$L(t_{10;x}; (2N)^{-\zeta}, \sigma\gamma) = x \text{ and } L(t_{10,x,y}; x, \sigma\gamma) = y.$$

We recall or define the following constants:

$$\begin{aligned}
\kappa &= \zeta/\gamma,\ \epsilon = (2N)^{-(\gamma-\zeta)/144}, \\
a_0 &= \frac{\zeta}{6},\ a_1 = \frac{\zeta}{8} \wedge \frac{1-\kappa}{4}, \\
t_{stoch} &= \frac{a_0}{\sigma}\log(2N),\ t_{early} = \frac{1.01\log(2N)}{\sigma(1-\gamma)-r}, \\
t_{mid} &= \frac{1}{\sigma(1-\gamma)}\log\frac{1-\epsilon}{\epsilon},\ t_{late} = \frac{1.02}{\sigma\gamma}\log(2N), \\
t'_{10;1-\epsilon^{24}} &= \begin{cases} t_{10;\epsilon^{24}} + t_{10,0.9\epsilon^{24},1-\epsilon^{24}}, & \text{if } \zeta > \gamma/7 \\ t_{10;1-\epsilon^{24}}, & \text{otherwise} \end{cases},
\end{aligned} \tag{4.1}$$

and stopping time

$$S_{X,Z,diff} = \inf\{t \geq 0 : X(t) \neq Z(t)\}$$

for processes X and Z. In the above, a_0 is picked so that

$$(2N)^{a_0-1} < \epsilon$$

for $\zeta < \gamma < 1$. And the case $\zeta > \gamma/7$ in the definition of $t'_{10;1-\epsilon^{24}}$ is for the case $(2N)^{-\zeta} > \epsilon^{24}$, so that X_{10} immediately enters its middle phase at $t = 0$. Finally, $t_{10,0.9\epsilon^{24},1-\epsilon^{24}}$ is the length of time for which we use event A_2 in Lemma 4.1 below. On event A_1^c defined in that lemma, X_{10} has reached $0.9\epsilon^{24}$ at time $t_{10;\epsilon^{24}}$, after which event A_2^c ensures X_{10} grows to levels slightly smaller than $1-\epsilon^{24}$ after another time period of length $t_{10,0.9\epsilon^{24},1-\epsilon^{24}}$. At $t_{10;\epsilon^{24}}$, the logistic curve $L(\cdot; (2N)^{-\zeta}, \sigma\gamma)$ reaches ϵ^{24}; we use $L(\cdot; (2N)^{-\zeta}, \sigma\gamma)$ to approximate X_{10}, and hence can only say that X_{10} reaches at least $0.9\epsilon^{24}$ at $t_{10;\epsilon^{24}}$. After $t_{10;\epsilon^{24}}$, X_{10} enters its middle phase and we use $L(\cdot; 0.9\epsilon^{24}, \sigma\gamma)$ (not $L(\cdot; \epsilon^{24}, \sigma\gamma)$) to approximate X_{10}, hence the time when $L(\cdot; (2N)^{-\zeta}, \sigma\gamma)$ is between $0.9\epsilon^{24}$ and ϵ^{24} is counted twice. We observe that

$$\begin{aligned}
t_{10;\epsilon^{24}} &= \begin{cases} \frac{1}{\sigma\gamma}\log\frac{(2N)^{\zeta}-1}{\frac{1}{\epsilon^{24}}-1}, & \text{if } \zeta > \gamma/7 \\ 0, & \text{otherwise} \end{cases} \\
t'_{10;1-\epsilon^{24}} &= \frac{1}{\sigma\gamma}\log\left[((2N)^{\zeta}-1)\left(\frac{1}{\epsilon^{24}}-1\right)\right] \\
&\quad + \frac{1}{\sigma\gamma}\log\frac{\frac{1}{0.9\epsilon^{24}}-1}{\frac{1}{\epsilon^{24}}-1}1_{\zeta>\gamma/7}
\end{aligned} \tag{4.2}$$

Lemma 4.1. *Let* $R_{01,11} = T_{11;1/(2N)} \wedge T_{01;(2N)^{-(1-\kappa)/2}}$. *We define*

$$\begin{aligned}
A_1 &= \{X_{10}(s) \leq 0.9L(s; (2N)^{-\zeta}, \sigma\gamma) \text{ for some } s \leq t_{10;\epsilon^{24}} \wedge R_{01,11}\} \\
A_2 &= \{X_{10}(s) < L(s - t_{10;\epsilon^{24}}; 0.9\epsilon^{24}1_{\zeta>\gamma/7} + (2N)^{-\zeta}1_{\zeta\leq\gamma/7}, \sigma\gamma) - \epsilon^{8} \\
&\qquad \text{for some } s \in [t_{10;\epsilon^{24}}, t'_{10;1-\epsilon^{24}} \wedge R_{01,11}]\}, \\
A_3 &= \{X_{01}(s) + X_{10}(s) \leq 1-\epsilon^{4} \text{ for some } s \in [t'_{10;1-\epsilon^{24}}, T_{11,1/(2N)})\}.
\end{aligned}$$

Then

$$
\begin{array}{ll}
(a) & \mathbb{P}(A_1) \le C_{\gamma,\sigma} N^{-(1-\zeta)/4} \\
(b) & \mathbb{P}(A_2 \cap A_1^c \cap \{t_{10;\epsilon^{24}} \le R_{01,11}\}) \le C_{\gamma,\sigma}\epsilon^8 \\
(c) & \mathbb{P}(A_3 \cap A_2^c \cap A_1^c \cap \{t'_{10;1-\epsilon^{24}} \le R_{01,11}\}) \le C_{\gamma,\sigma} N^{-1/2}.
\end{array}
$$

Consequently,

$$\mathbb{P}((A_3 \cup A_2 \cup A_1) \cap \{t'_{10;1-\epsilon^{24}} \le R_{01,11}\}) \le C_{\gamma,\sigma}\epsilon^8.$$

Proof. The proof essentially consists of identifying the constants required for straightforward applications of Lemmas 5.1 to 5.3 for each of the three phases of X_{10}. We only prove the more complicated case of $\zeta > \gamma/7$. The proof for the case $\zeta \le \gamma/7$ involves only events A_2 and A_3, which correspond to the middle and late phases, respectively, and follows by a similar argument.

(a) Early Phase. Before the stopping time $R_{01,11}$, the jump rates of X_{10} satisfy

$$
\begin{aligned}
r_{10}^+ &\ge NX_{10}[(1+\sigma\gamma+r)(1-X_{10}) - 1.1\sigma(2N)^{-(1-\kappa)/2}], \\
r_{10}^- &\le NX_{10}[(1-\sigma\gamma+r)(1-X_{10}) + 1.1\sigma(2N)^{-(1-\kappa)/2}].
\end{aligned}
$$

We take $\hat{\xi} = X_{10}$, $\alpha = 1+r$, $\beta = \sigma\gamma$, $\delta_0 = 1.1\sigma(2N)^{-(1-\kappa)/2}$, $\delta_1 = \epsilon^{24}$, $\delta_2 = (1-\zeta)/4$, $x = \zeta$, $Y(t) = (2N)^{-\zeta} + \int_0^t Y(s)(\sigma\gamma(1-Y(s)) - 1.1\sigma(2N)^{-(1-\kappa)/2})\, ds$, and $u_0 = \inf\{t : Y(t) = \delta_1\} > t_{10;\epsilon^{24}}$ in Lemma 5.1 to obtain

$$\mathbb{P}\left(X_{10}(s) < 0.99Y(s) \text{ for some } s \le t_{10;\epsilon^{24}} \wedge R_{01,11}\right) \le C_{\gamma,\sigma} N^{-(1-\zeta)/4}.$$

Prior to u_0, the deterministic function Y is sandwiched between $L(\cdot;(2N)^{-\zeta}, \sigma\gamma - 1.2\sigma(2N)^{-(1-\kappa)/2})$ and $L(\cdot;(2N)^{-\zeta},\sigma\gamma)$. Since

$$L(t;(2N)^{-\zeta},\sigma\gamma) - L(t;(2N)^{-\zeta},\sigma\gamma - v) \le (1-e^{-vt})L(t;(2N)^{-\zeta},\sigma\gamma)$$

for $v \le \sigma\gamma$, we can take $v = 1.2\sigma(2N)^{-(1-\kappa)/2}$ in the above and obtain

$$
\begin{aligned}
Y(t) &\ge L(t;(2N)^{-\zeta},\sigma\gamma - 1.2\sigma(2N)^{-(1-\kappa)/2}) \\
&\ge e^{-1.2\sigma(2N)^{-(1-\kappa)/2}t} L(t;(2N)^{-\zeta},\sigma\gamma) \\
&\ge 0.99L(t;(2N)^{-\zeta},\sigma\gamma)
\end{aligned}
$$

for $t = \mathcal{O}(\log N)$. Hence

$$
\begin{aligned}
&\mathbb{P}\left(X_{10}(s) < 0.99^2 L(t;(2N)^{-\zeta},\sigma\gamma) \text{ for some } s \le t_{10;\epsilon^{24}} \wedge R_{01,11}\right) \\
&\quad \le \mathbb{P}\left(X_{10}(s) < 0.99Y(s) \text{ for some } s \le t_{10;\epsilon^{24}} \wedge R_{01,11}\right) \le C_{\gamma,\sigma} N^{-(1-\zeta)/4}
\end{aligned}
$$

and (a) follows.

(b) Middle Phase. Before $R_{01,11}$, $X_{11} = 0$. Using the jump rates of X_{10} in (2.1), we can write

$$
\begin{aligned}
X_{10}(t \wedge R_{01,11}) &= b_0 + M_{10}(t \wedge R_{01,11}) \\
&\quad + \int_{u_1}^{t \wedge R_{01,11}} X_{10}(s)[\sigma\gamma(1-X_{10}(s)) - (\sigma+r)X_{01}(s)]\, ds,
\end{aligned}
$$

where $M_{10}(\cdot \wedge R_{01,11})$ is a martingale with maximum jump size $1/(2N)$ and quadratic variation

$$\langle M_{10}\rangle(t \wedge R_{01,11}) = \frac{1+r}{2N}\int_{u_1}^{t \wedge R_{01,11}} X_{10}(s)(1-X_{10}(s))\, ds.$$

We apply Lemma 5.2 with $b_0 = X_{10}(t_{10;\epsilon^{24}})$, $u_1 = t_{10;\epsilon^{24}}$, $u_2 = t'_{10;1-\epsilon^{24}}$, $b_1 = 1 - L(t_{10,0.9\epsilon^{24},\epsilon^{24}}, b_0, \sigma\gamma)$, $\delta_1 = (1-\kappa)/2$, $\delta_2 = \infty$, $\delta_3 = \gamma(1-\kappa)/18 = (\gamma-\zeta)/18$, $\epsilon_0 = \epsilon_1 = (\gamma-\zeta)/6 = -\log_{2N}\epsilon^{24}$, $\epsilon_2(t) = -(\sigma+r)X_{01}(t)$, $\epsilon_3(t) = \epsilon_4(t) = 0$, $\beta = \sigma\gamma$, $\xi = X_{10}$, $Y = L(\cdot; b_0, \sigma\gamma)$, $T = R_{01,11}$ and $D_1 = A_1^c$, and obtain

$$\begin{aligned}&\mathbb{P}(\{|X_{10}(s) - L(s - t_{10;\epsilon^{24}}; X_{10}(t_{10;\epsilon^{24}}), \sigma\gamma))| > \epsilon^8 \text{ for some}\\ &\quad s \in [t_{10;\epsilon^{24}}, t'_{10;1-\epsilon^{24}} \wedge R_{01,11}]\} \cap A_1^c \cap \{t_{10;\epsilon^{24}} \le R_{01,11}\}) \le \epsilon^8,\end{aligned}$$

where $-\log_{2N}\epsilon^8 = (\gamma-\zeta)/18 < (\delta_1 - \epsilon_0 - \epsilon_1)/3$. Now for paths in $A_1^c \cap \{t_{10;\epsilon^{24}} \le R_{01,11}\}$, we have $X_{10}(t_{10;\epsilon^{24}}) \ge 0.9\epsilon^{24}$ and hence

$$L(s - t_{10;\epsilon^{24}}; X_{10}(t_{10;\epsilon^{24}}), \sigma\gamma) \ge L(s - t_{10;\epsilon^{24}}; 0.9\epsilon^{24}, \sigma\gamma).$$

The desired conclusion in (b) follows.

(c) Late Phase. On $A_1^c \cap A_2^c \cap \{t'_{10;1-\epsilon^{24}} \le R_{01,11}\}$, we have

$$\begin{aligned} X_{10}(t'_{10;1-\epsilon^{24}}) &> L(t'_{10;1-\epsilon^{24}} - t_{10;\epsilon^{24}}; 0.9\epsilon^{24}, \sigma\gamma) - \epsilon^8 \\ &= 1 - \epsilon^{24} - \epsilon^8. \end{aligned}$$

Therefore $X_{00}(t'_{10;1-\epsilon^{24}}) < 2\epsilon^8$. Before $T_{11,1/(2N)}$, $X_{11} = 0$, and the jump rates of X_{00} satisfy

$$\begin{aligned} r_{00}^{+} &\le N(1 - \sigma\gamma + r)X_{00}(1 - X_{00}), \\ r_{00}^{-} &\ge N(1 + \sigma\gamma + r)X_{00}(1 - X_{00}). \end{aligned}$$

With $\alpha = 1 + r$, $\beta = \sigma\gamma$, $K = 0$, $c_4 = 2$, $x = (\gamma-\zeta)/6$ and $c_5 = 1/2$, Lemma 5.3 implies

$$\mathbb{P}\left(\left\{\sup_{t \ge t'_{10;1-\epsilon^{24}}} X_{00}(t) \ge \epsilon^4\right\} \cap A_1^c \cap A_2^c \cap \{t'_{10;1-\epsilon^{24}} \le R_{01,11}\}\right) \le C_{\gamma,\sigma} N^{-1/2},$$

which implies the desired conclusion in (c). □

For the remainder of this section, we define the following events

$$\begin{aligned} A_{41} &= \{X_{01}(s) \ge (2N)^{a_0+a_1-1} \text{ for some } s \le t_{stoch} \wedge T_{10;(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}\} \\ A_{42} &= \{X_{01}(t_{stoch}) \in [(2N)^{-1}, (2N)^{a_0-a_1-1}]\} \\ A_4 &= A_{41} \cup A_{42} \cup E_1^c \\ B_4 &= \{t_{stoch} \le T_{10;(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}\} \\ A_{51} &= \{X_{01}(s) \ge (2N)^{-(1-\kappa)/2} \text{ for some } s \in [t_{stoch}, t'_{10;1-\epsilon^{24}} \wedge T_{11;1/(2N)}]\} \\ A_{52} &= \{T_{01;\epsilon} \wedge T_{11;1/(2N)} \ge t_{stoch} + t_{early}\}. \end{aligned}$$

Event B_4 ensures that X_{10} has not become too large by t_{stoch} nor has there been a recombination event leading to the birth of a type 11 individual. It has a high probability. On B_4, we can estimate the probability of bad events (A_{41}, A_{42}, A_{51} and A_{52}) related to X_{01}. Event A_4^c ensures that if type 01 remains positive at t_{stoch} (event E_1), then X_{01} is neither too large (event A_{41}^c) nor too small (event A_{42}^c). Once we establish that $X_{01} \in [(2N)^{a_0-a_1-1}, (2N)^{a_0+a_1-1}]$ at t_{stoch}, event A_{52} ensures that X_{01} grows deterministically to ϵ by $t_{stoch} + t_{early}$. Event A_{51} is an upper bound for X_{01} for later use in the proof of Proposition 3.1(a-b).

Lemma 4.2. *Recall* $t_{stoch} = \frac{a_0}{\sigma}\log(2N)$, $a_0 = \frac{\zeta}{6}$, *and* $a_1 = \frac{\zeta}{8} \wedge \frac{1-\kappa}{4}$. *Let*

$$\begin{aligned} c_1 &= (2N)^{a_0+a_1-1} \\ c_2 &= (2N)^{a_0+a_1}((2N)^{a_0+a_1-1} + (2N)^{-\zeta/3}). \end{aligned}$$

We have

$$\begin{aligned} &(a) \quad \mathbb{P}(B_4^c) \le C_{\gamma,\sigma}(N^{-(1-\zeta)/4} + N^{-\zeta/4}) \\ &(b) \quad \mathbb{P}(A_{41}) \le 2c_2 t_{stoch} + C_{\gamma,\sigma}N^{-a_1} \\ &\qquad \mathbb{P}(A_{42} \cap A_{41}^c \cap B_4) \le C_{\gamma,\sigma}N^{-a_1} + 2c_2 t_{stoch} \\ &\qquad \left|\mathbb{P}(E_1^c \cap B_4) - \frac{1-\sigma+r}{1+\sigma+r}\right| \le 8c_2 t_{stoch} + C_{\gamma,\sigma}N^{-a_1} \\ &(c) \quad \mathbb{P}\left(A_{51} \cap A_4^c\right) \le C_{\gamma,\sigma}(N^{-(a_0-a_1)/4} + N^{-(1-\zeta)/4} + N^{-\zeta/4}). \\ &(d) \quad \mathbb{P}(A_{52} \cap A_4^c \cap B_4) \le C_{\gamma,\sigma}N^{-(a_0-a_1)/4}. \end{aligned}$$

Proof. **(a) The Event B_4.** We first show

$$B_4^c = \{t_{stoch} > T_{10;(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}\}$$

has a small probability. This gives an upper bound for X_{10} during the stochastic phase of X_{01}. As a result, it is unlikely for type 01 and 10 individuals to recombine to produce a type 11 individual. Let

$$F_1 = \{T_{10;(2N)^{-\zeta/3}} < t_{stoch} \wedge T_{11;1/(2N)}\}.$$

Before $T_{11;1/(2N)}$, the jump rates of X_{10} satisfy

$$\begin{aligned} r_{10}^+ &\le N(1+\sigma\gamma+r)X_{10}(1-X_{10}) \\ r_{10}^- &\ge N(1-\sigma\gamma+r)X_{10}(1-X_{10}). \end{aligned}$$

We take $\check{\xi} = X_{10}$, $\alpha = 1+r$, $\beta = \sigma\gamma$, $\delta_0 = 0$, $\delta_1 = 0.9(2N)^{-\zeta/3}$, $\delta_2 = (1-\zeta)/4$, $x = \zeta$ and $Y(t) = L(t;(2N)^{-\zeta},\sigma\gamma)$ in Lemma 5.1 to obtain

$$\mathbb{P}\left(X_{10}(s) \ge (2N)^{-\zeta/3} \text{ for some } s \le t_{10;0.9(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}\right) \le C_{\gamma,\sigma}N^{-(1-\zeta)/4}.$$

By the choice of a_0, $t_{stoch} = \frac{\zeta}{6\sigma}\log(2N) < \frac{\zeta}{6\sigma\gamma}\log(2N) = \frac{1}{\sigma\gamma}\log(2N)^{\zeta/6} < t_{10;0.9(2N)^{-\zeta/3}} = \frac{1}{\sigma\gamma}\log\frac{(2N)^{\zeta}-1}{0.9(2N)^{\zeta/3}-1}$, therefore

$$\mathbb{P}(F_1) \le \mathbb{P}(T_{10;(2N)^{-\zeta/3}} \le t_{10;0.9(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}) \le C_{\gamma,\sigma}N^{-(1-\zeta)/4}.$$

We observe that

$$\begin{aligned} &B_4^c \cap F_1^c \\ &\quad = \{t_{stoch} > T_{10;(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}\} \cap \{T_{10;(2N)^{-\zeta/3}} \ge t_{stoch} \wedge T_{11;1/(2N)}\} \\ &\quad \subset \{t_{stoch} \wedge T_{10;(2N)^{-\zeta/3}} > T_{11;1/(2N)}\}. \end{aligned}$$

Before $t_{stoch} \wedge T_{10;(2N)^{-\zeta/3}}$, the rate of recombination events between type 01 and 10 individuals is at most $4rNX_{01}X_{10} \le 4rN(2N)^{-\zeta/3} \le CN^{-\zeta/3}$. Hence the total number of recombination events between type 01 and 10 individuals before $t_{stoch} \wedge T_{10;(2N)^{-\zeta/3}}$

is dominated by a Poisson random variable with mean $C_{\gamma,\sigma}N^{-\zeta/3}\log N$ which implies that

$$\mathbb{P}(B_4^c \cap F_1^c) \leq C_{\gamma,\sigma}N^{-\zeta/4}$$

for large enough N, so that

$$\mathbb{P}(B_4^c) \leq C_{\gamma,\sigma}(N^{-(1-\zeta)/4} + N^{-\zeta/4}),$$

as required by (a).

(b) Stochastic Phase. We couple X_{01} to a branching process η, defined below, and establish the required estimates for η, which easily implies corresponding estimates for X_{01}. We define $R_{10,11} = T_{10;(2N)^{-\zeta/3}} \wedge T_{11;1/(2N)}$ and

$$A_6 = \{S_{X_{01},\eta,diff} < t_{stoch} \wedge T_{01;c_1} \wedge R_{10,11}\}.$$

Before $T_{11;1/(2N)}$, the jump rates of X_{01} are as follows:

$$\begin{aligned} r_{01}^+ &= NX_{01}[(1+\sigma+r)(1-X_{01}) - (\sigma\gamma + r)X_{10}], \\ r_{01}^- &= NX_{01}[(1-\sigma+r)(1-X_{01}) + (\sigma\gamma + r)X_{10}]. \end{aligned}$$

We define η to be a jump process with $\eta(0) = 1/(2N)$, jump size $1/(2N)$ and jump rates as follows:

$$\begin{aligned} r_\eta^+ &= N\eta(1+\sigma+r), \\ r_\eta^- &= N\eta(1-\sigma+r). \end{aligned}$$

Then prior to $S_{X_{01},\eta,diff} \wedge T_{01;c_1} \wedge R_{10,11}$, we have $|r_{01}^+ - r_\eta^+| \leq c_2$ and $|r_{01}^- - r_\eta^-| \leq c_2$. Therefore $|X_{01} - \eta|$ is a jump process with initial value 0, jump size $1/(2N)$ and jump rates at most $2c_2$, and we can estimate the probability of $|X_{01} - \eta|$ becoming nonzero before t_{stoch}:

$$\mathbb{P}(A_6) \leq 2c_2 t_{stoch}, \tag{4.3}$$

which $\downarrow 0$ as $N \to \infty$. Since η is a branching process, Lemma 6.1(a) implies

$$\begin{aligned} \mathbb{P}\left(\sup_{s \leq t_{stoch}} \eta(s) \geq c_1 = (2N)^{a_0+a_1-1}\right) &\leq C_{\gamma,\sigma}N^{-a_1} \\ \mathbb{P}((2N)^{-1} \leq \eta(t_{stoch}) \leq (2N)^{a_0-a_1-1}) &\leq C_{\gamma,\sigma}N^{-a_1} \\ \left|\mathbb{P}(\eta(t_{stoch}) = 0) - \frac{1-\sigma+r}{1+\sigma+r}\right| &\leq \frac{1-\sigma+r}{1+\sigma+r}e^{-2\sigma t_{stoch}} \leq (2N)^{-a_0}. \end{aligned}$$

Using (4.3), we can replace η in the above three estimates by X_{01} if we allow an additional error term. In particular,

$$\begin{aligned} \mathbb{P}(A_{41}) &= \mathbb{P}(T_{01;c_1} \leq t_{stoch} \wedge R_{10,11}) \\ &\leq \mathbb{P}(\{T_{01;c_1} \leq t_{stoch} \wedge R_{10,11}\} \cap A_6^c) + \mathbb{P}(A_6) \\ &= \mathbb{P}(T_{01;c_1} \leq t_{stoch} \wedge R_{10,11} \wedge S_{X_{01},\eta,diff}) + \mathbb{P}(A_6) \\ &\leq \mathbb{P}\left(\sup_{s \leq t_{stoch} \wedge R_{10,11}} \eta(s) \geq c_1\right) + \mathbb{P}(A_6) \\ &\leq 2c_2 t_{stoch} + C_{\gamma,\sigma}N^{-a_1}. \end{aligned}$$

Similarly, we can obtain the second statement of (b):

$$\begin{aligned}&\mathbb{P}(A_{42}\cap A_{41}^c\cap B_4)\le \mathbb{P}(A_{42}\cap A_{41}^c\cap B_4\cap A_6^c)+\mathbb{P}(A_6)\\ &\le\quad \mathbb{P}(X_{01}(t_{stoch})\in[(2N)^{-1},(2N)^{a_0-a_1-1}],S_{X_{01},\eta,diff}\ge t_{stoch})+\mathbb{P}(A_6)\\ &\le\quad 2c_2t_{stoch}+C_{\gamma,\sigma}N^{-a_1}.\end{aligned}$$

Finally, for the third statement of (b), we have

$$\begin{aligned}&|\mathbb{P}(\{X_{01}(t_{stoch})=0\}\cap A_6^c\cap A_{41}^c\cap B_4)-\frac{1-\sigma+r}{1+\sigma+r}|\\ &=\quad |\mathbb{P}(\{\eta(t_{stoch})=0\}\cap A_6^c\cap A_{41}^c\cap B_4)-\frac{1-\sigma+r}{1+\sigma+r}|\\ &\le\quad |\mathbb{P}(\{\eta(t_{stoch})=0\}\cap A_6^c\cap A_{41}^c\cap B_4)-\mathbb{P}(\eta(t_{stoch})=0)|+(2N)^{-a_0}\\ &\le\quad \mathbb{P}(A_6\cup A_{41}\cup B_4^c)+(2N)^{-a_0}\\ &\le\quad 4c_2t_{stoch}+C_{\gamma,\sigma}(N^{-a_1}+N^{-(1-\zeta)/4}+N^{-\zeta/4}+N^{-a_0}).\end{aligned}$$

Hence by the estimates we have obtained in this part,

$$\begin{aligned}&\left|\mathbb{P}(E_1^c\cap B_4)-\frac{1-\sigma+r}{1+\sigma+r}\right|\\ &\quad=\left|\mathbb{P}(E_1^c\cap B_4\cap(A_6\cup A_{41}))+\mathbb{P}(E_1^c\cap B_4\cap A_6^c\cap A_{41}^c)-\frac{1-\sigma+r}{1+\sigma+r}\right|\\ &\quad\le\mathbb{P}(A_6\cup A_{41})+\left|\mathbb{P}(E_1^c\cap B_4\cap A_6^c\cap A_{41}^c)-\frac{1-\sigma+r}{1+\sigma+r}\right|\\ &\quad\le 8c_2t_{stoch}+C_{\gamma,\sigma}(N^{-a_1}+N^{-(1-\zeta)/4}+N^{-\zeta/4}+N^{-a_0}).\end{aligned}$$

which implies the third statement in (b).

(c) Early Phase (Upper Bound for X_{01}). We apply Lemma 5.1 with the process ξ dominating X_{01}, hence giving us an upper bound for X_{01} during the early and middle phases of X_{10}. Before $T_{11;1/(2N)}$, the jump rates of X_{01} satisfy

$$\begin{aligned}r_{01}^+&\le N(1+\sigma+r)X_{01}(1-X_{01}),\\ r_{01}^-&\ge N(1-\sigma+r)X_{01}(1-X_{01}).\end{aligned}$$

We take $\check{\xi}=X_{01}$, $\alpha=1+r$, $\beta=\sigma$, $\delta_0=0$, $\delta_1=0.9(2N)^{-(1-\kappa)/2}$, $\delta_2=(a_0-a_1)/4$, $Y(t)=L(t;X_{01}(t_{stoch}),\sigma)$, and $u_0=R_2=\inf\{t\ge 0: L(t;X_{01}(t_{stoch}),\sigma)\ge\delta_1\}$ in Lemma 5.1 to obtain

$$\begin{aligned}&\mathbb{P}(\{X_{01}(t_{stoch}+s)\ge 1.01L(s;X_{01}(t_{stoch}),\sigma)\text{ for some}\\ &\qquad s\le(R_2\wedge T_{11;1/(2N)})-t_{stoch}]\}\cap A_4^c\cap B_4)\le C_{\gamma,\sigma}N^{-(a_0-a_1)/4}.\end{aligned}$$

By the choice of a_1, we have the following in turn:

$$\begin{aligned}a_1+\frac{1-\kappa}{2}+\frac{1-\kappa}{6}\quad&<\quad 1-\kappa=1-\frac{\zeta}{\gamma},\\ (1-a_1)\log(2N)+\log(2N)^{-(1-\kappa)/2}+\frac{1}{\gamma}\log\epsilon^{24}\quad&>\quad\frac{1}{\gamma}\log((2N)^\zeta-1),\end{aligned}$$

$$\begin{aligned}&\log((2N)^{1-a_1}-(2N)^{a_0})-\log\left(\frac{1}{0.9(2N)^{-(1-\kappa)/2}}-1\right)-\frac{1}{\gamma}\log\left(\frac{1}{\epsilon^{24}}-1\right)\\ &\qquad\qquad>\frac{1}{\gamma}\log((2N)^\zeta-1)+\frac{1}{\gamma}\log\frac{1}{0.9},\end{aligned}$$

$$\begin{aligned}&\log((2N)^{1-a_1}-(2N)^{a_0})-\log\left(\frac{1}{0.9(2N)^{-(1-\kappa)/2}}-1\right)\\&\geq\frac{1}{\gamma}\left\{\log\left[((2N)^{\zeta}-1)\left(\frac{1}{\epsilon^{24}}-1\right)\right]+\log\frac{\frac{1}{0.9\epsilon^{24}}-1}{\frac{1}{\epsilon^{24}}-1}\right\}\\&\geq\sigma t'_{10;1-\epsilon^{24}},\end{aligned}$$

where $t'_{10;1-\epsilon^{24}}$ is defined in (4.2). Therefore, on $A_4^c \cap B_4 \subset \{X_{01}(t_{stoch}) \in ((2N)^{a_0-a_1-1}, (2N)^{a_0+a_1-1})\}$, we have

$$\begin{aligned}&t_{stoch}+R_2\\&\geq\quad\frac{1}{\sigma}\left[a_0\log(2N)+\log((2N)^{1-a_0-a_1}-1)-\log\left(\frac{1}{0.9(2N)^{-(1-\kappa)/2}}-1\right)\right]\\&\geq\quad t'_{10;1-\epsilon^{24}}.\end{aligned}$$

Hence if

$$X_{01}(t_{stoch}) \in ((2N)^{a_0-a_1-1}, (2N)^{a_0+a_1-1}),$$

then $L(t'_{10;1-\epsilon^{24}} - t_{stoch}; X_{01}(t_{stoch}), \sigma) \leq L(R_2; X_{01}(t_{stoch}), \sigma) = 0.9(2N)^{-(1-\kappa)/2}$, which implies

$$\mathbb{P}\left(A_{51} \cap A_4^c \cap B_4\right) \leq C_{\gamma,\sigma} N^{-(a_0-a_1)/4}.$$

We combine the above inequality and part (a) of this lemma to obtain the desired inequality in (c).

(d) Early Phase (Lower Bound). Before $T_{11;1/(2N)}$, the jump rates of X_{01} satisfy

$$\begin{aligned}r_{01}^+ &\geq N(1+\sigma(1-\gamma))X_{01}(1-X_{01})\\ r_{01}^- &\leq N(1-\sigma(1-\gamma)+2r)X_{01}(1-X_{01}).\end{aligned}$$

We take $\hat{\xi}$ to be X_{01} shifted forward in time by t_{stoch}, $\alpha = 1+r$, $\beta = \sigma(1-\gamma)-r$, $\delta_0 = 0$, $\delta_1 = 1.01\epsilon$, $\delta_2 = (a_0-a_1)/4$, $x = 1-a_0+a_1$, $Y(t) = L(t; (2N)^{a_0-a_1-1}, \sigma(1-\gamma)-r)$ and $u_0 = \inf\{t : Y(t) = 1.01\epsilon\}$ in Lemma 5.1 to obtain

$$\begin{aligned}&\mathbb{P}(\{X_{01}(t_{stoch}+s) < 1.005L(s; N^{a_0-a_1-1}, \sigma(1-\gamma)-r)\\&\qquad\text{for some } s \leq u_0 \wedge (T_{11;1/(2N)} - t_{stoch})\} \cap A_4^c \cap B_4) \leq C_{\gamma,\sigma} N^{-(a_0-a_1)/4}.\end{aligned}$$

Since $u_0 \leq t_{early} = \frac{1.01}{\sigma(1-\gamma)-r}\log(2N)$, the conclusion in (d) follows. □

Proof of Proposition 3.1(a-b). Part (a) of the proposition follows from part (a) and the third statement of part (b) in Lemma 4.2. For part (b), we observe that event G_1 is an intersection of two events. In what follows, we will first show that on E_1,

$$\{T_{01;\epsilon} \leq T_{11;1/(2N)} \wedge (t_{stoch}+t_{early})\} \cap \{X_{10}(T_{01;\epsilon}) \geq 1-\epsilon-\epsilon^4\}^c$$

is unlikely to occur, then show neither is

$$\{T_{01;\epsilon} \leq T_{11;1/(2N)} \wedge (t_{stoch}+t_{early})\}^c$$

likely. We define

$$F_2 \quad = \quad \left\{T_{11;1/(2N)} \leq T_{01;\epsilon} \wedge ((t_{stoch}+t_{early}) \vee t'_{10;1-\epsilon^{24}})\right\}.$$

Before $T_{01;\epsilon} \wedge ((t_{stoch}+t_{early}) \vee t'_{10;1-\epsilon^{24}})$, the rate of recombination events between type 01 and 10 individuals is at most $4rNX_{01}X_{10} \leq 4rN\epsilon \leq C\epsilon$. Hence the total number of recombination events between type 01 and 10 individuals before $T_{01;\epsilon}\wedge((t_{stoch}+t_{early})\vee t'_{10;1-\epsilon^{24}})$ is dominated by a Poisson random variable with mean $C_{\gamma,\sigma}\epsilon\log N$. Therefore

$$\mathbb{P}(F_2)\le C_{\gamma,\sigma}\epsilon^{7/8}. \tag{4.4}$$

Now we try to estimate the probability of

$$F_3=\{t'_{10;1-\epsilon^{24}}\le T_{11;1/(2N)}\wedge T_{01;(2N)^{-(1-\kappa)/2}}\},$$

so that we can drop this event in the conclusion of Lemma 4.1. We observe that

$$\begin{aligned}&F_2^c\cap F_3^c\\ &\quad\subset\ \left\{T_{11;1/(2N)}>t'_{10;1-\epsilon^{24}}\wedge T_{01;\epsilon}\right\}\cap\{t'_{10;1-\epsilon^{24}}>T_{11;1/(2N)}\wedge T_{01;(2N)^{-(1-\kappa)/2}}\}\\ &\quad\subset\ \{t'_{10;1-\epsilon^{24}}\wedge T_{11;1/(2N)}>T_{01;(2N)^{-(1-\kappa)/2}}\}.\end{aligned}$$

Since $F_2^c\cap F_3^c\cap A_4^c\cap B_4\subset A_{51}\cap A_4^c\cap B_4$ and $A_4^c=A_{41}^c\cap A_{42}^c\cap E_1$, we have

$$\begin{aligned}&\mathbb{P}(F_2^c\cap F_3^c\cap A_{41}^c\cap A_{42}^c\cap E_1\cap B_4)\\ &\quad=\ \mathbb{P}(F_2^c\cap F_3^c\cap A_4^c\cap B_4)\le\mathbb{P}(A_{51}\cap A_4^c\cap B_4)\\ &\quad\le\ C_{\gamma,\sigma}(N^{-(a_0-a_1)/4}+N^{-(1-\zeta)/4}+N^{-\zeta/4})\end{aligned}$$

by Lemma 4.2(c). Combining the above with (4.4) and Lemma 4.2(a-b) yields

$$\begin{aligned}&\mathbb{P}(F_3^c\cap E_1)\\ &\quad\le\ 4c_2t_{stoch}+C_{\gamma,\sigma}(\epsilon^{7/8}+N^{-a_1}+N^{-(1-\zeta)/4}+N^{-\zeta/4}+N^{-(a_0-a_1)/4})\\ &\quad\le\ C_{\gamma,\sigma}N^{-\delta}\end{aligned}$$

for some δ, whose value may subsequently change.

We observe that on $\{T_{01;\epsilon}\le T_{11;1/(2N)}\wedge(t_{stoch}+t_{early})\}\cap F_3$, we have $t'_{10;1-\epsilon^{24}}\le T_{01;(2N)^{-(1-\kappa)/2}}\le T_{01;\epsilon}\le T_{11;1/(2N)}$, therefore $X_{10}(T_{01;\epsilon})<1-\epsilon-\epsilon^4$ implies $X_{01}(T_{01;\epsilon})+X_{10}(T_{01;\epsilon})<1-\epsilon^4$, i.e.

$$\begin{aligned}&\mathbb{P}(\{T_{01;\epsilon}\le T_{11;1/(2N)}\wedge(t_{stoch}+t_{early})\}\cap\{X_{10}(T_{01;\epsilon})\ge 1-\epsilon-\epsilon^4\}^c\cap F_3)\\ &\qquad\qquad\le\mathbb{P}((A_3\cup A_2\cup A_1)\cap F_3)\le C_{\gamma,\sigma}\epsilon^8,\end{aligned}$$

by Lemma 4.1. We combine the above two estimates to obtain

$$\begin{aligned}&\mathbb{P}(\{T_{01;\epsilon}\le T_{11;1/(2N)}\wedge(t_{stoch}+t_{early})\}\cap\{X_{10}(T_{01;\epsilon})\ge 1-\epsilon-\epsilon^4\}^c\cap E_1)\\ &\qquad\qquad\le C_{\gamma,\sigma}N^{-\delta}.\end{aligned} \tag{4.5}$$

Now we try to estimate the probability of $\{T_{01;\epsilon}>T_{11;1/(2N)}\wedge(t_{stoch}+t_{early})\}$. Parts (a), (b) and (d) of Lemma 4.2 imply

$$\mathbb{P}(A_{52}\cap E_1)\le\mathbb{P}(A_{52}\cap A_{41}^c\cap A_{42}^c\cap E_1)+\mathbb{P}(A_{41}\cup A_{42})\le C_{\gamma,\sigma}N^{-\delta}.$$

We also have

$$\begin{aligned}&\mathbb{P}(\{T_{01;\epsilon}>T_{11;1/(2N)}\wedge(t_{stoch}+t_{early})\}\cap A_{52}^c\cap E_1)\\ &\quad=\ \mathbb{P}(\{T_{01;\epsilon}>(t_{stoch}+t_{early})\wedge T_{11;1/(2N)}\}\\ &\qquad\qquad\cap\{t_{stoch}+t_{early}>T_{01;\epsilon}\wedge T_{11;1/(2N)}\}\cap E_1)\\ &\quad=\ \mathbb{P}(\{T_{01;\epsilon}\wedge(t_{stoch}+t_{early})>T_{11;1/(2N)}\}\cap E_1)\\ &\quad\le\ \mathbb{P}(F_2)\le C_{\gamma,\sigma}\epsilon^{7/8}.\end{aligned}$$

by (4.4). The two estimates above imply

$$\mathbb{P}(\{T_{01;\epsilon}>T_{11;1/(2N)}\wedge(t_{stoch}+t_{early})\}\cap E_1)\le C_{\gamma,\sigma}N^{-\delta}. \tag{4.6}$$

We combine (4.5) and (4.6) to obtain the desired result in (b). □

Proof of Proposition 3.1(c-e). In (c), we deal with the middle phase of X_{01} and show that it is well approximated by Z_{01}. Parts (d) and (e) deal with the late phase of X_{01}. Here we couple $X_{11}+X_{01}$ to a branching process η, defined below, and show that η does not stray too far away from 1 (part (d)) and hits 1 by T_∞ (part (e)). Recall that

$$Z_{01}(T_{01;\epsilon}+t) = L(t;\epsilon,\sigma(1-\gamma))$$

for $t \in [T_{01;\epsilon}, T_{Z_{01};1-\epsilon}]$, and

$$T_{Z_{01};1-\epsilon} = T_{01;\epsilon} + \frac{1}{\sigma(1-\gamma)} \log\frac{1-\epsilon}{\epsilon}.$$

We work on $t \geq T_{01;\epsilon}$ throughout this proof. On $G_1 \cap E_1$, we have

$$\begin{aligned} X_{10}(T_{01;\epsilon}) &\geq 1-\epsilon-\epsilon^4 \\ X_{01}(T_{01;\epsilon}) &= \epsilon \\ X_{00}(T_{01;\epsilon}) &\leq \epsilon^4. \end{aligned}$$

We can then write down the following equation using the jump rates of X_{01} in (2.1):

$$\begin{aligned} X_{01}(t) &= \epsilon + M_{01}(t) + \int_{T_{01;\epsilon}}^t X_{01}(s)[\sigma(1-\gamma)(1-X_{01}(s)) \\ &\quad -\sigma X_{11}(s) + \sigma\gamma X_{00}(s)] + r(X_{11}(s)X_{00}(s) - X_{01}(s)X_{10}(s))\, ds, \end{aligned}$$

where M_{01} is a martingale with maximum jump size $1/(2N)$ and quadratic variation

$$\langle M_{01}\rangle(t) = \frac{1}{2N}\int_{T_{01;\epsilon}}^t (1+r)X_{01}(s)(1-X_{01}(s)) + rX_{11}(s)X_{00}(s)\, ds.$$

We use Lemma 5.2 with $\beta = \sigma(1-\gamma)$, $u_1 = 0$, $u_2 = \frac{1}{\sigma(1-\gamma)}\log\frac{1-\epsilon}{\epsilon}$, $\delta_1 = -\log_{2N}\epsilon^3$, $\delta_2 = \infty$, $\delta_3 = -\log_{2N}\epsilon^{1/3}$, $b_0 = b_1 = \epsilon$, $\epsilon_0 = \epsilon_1 = -\log_{2N}\epsilon$, $T = T_{11;\theta_{11}} \wedge T_{00;\epsilon^3}$, $\epsilon_2(t) = -\sigma X_{11}(t) + \sigma\gamma X_{00}(t)$, $\epsilon_3(t) = r(X_{11}(t)X_{00}(t) - X_{10}(t)X_{01}(t))$, $\epsilon_4(t) = X_{11}(t)X_{00}(t)$, $Y(t) = Z_{01}(T_{01;\epsilon}+t)$, and $D_1 = G_1\cap E_1$ to obtain

$$\begin{aligned} \mathbb{P}\Big(&|X_{01}(s,\omega) - Z_{01}(s,\omega)| > \epsilon^{1/3} \text{ for some } \omega \in G_1\cap E_1, \\ &s \in [T_{01;\epsilon}, T_{Z_{01};1-\epsilon} \wedge T_{11;\theta_{11}} \wedge T_{00;\epsilon^3}\Big) \leq \epsilon^{1/3}. \end{aligned} \tag{4.7}$$

The jump rates of X_{00} satisfy

$$\begin{aligned} r_{00}^+ &\leq N[(1-\sigma\gamma+r)X_{00}(1-X_{00}) + 2rX_{10}X_{01}], \\ r_{00}^- &\geq N(1+\sigma\gamma+r)X_{00}(1-X_{00}). \end{aligned}$$

On $G_1 \cap E_1$, we have $X_{00}(T_{01;\epsilon}) \leq \epsilon^4$. Therefore by Lemma 5.3,

$$\mathbb{P}\left(\left\{\sup_{s\in[T_{01;\epsilon},T_{Z_{01};1-\epsilon}]} X_{00}(s) \geq \epsilon^3\right\} \cap G_1 \cap E_1\right) \leq C_{\gamma,\sigma}N^{(\gamma-\zeta)/144-1/2} \leq C_{\gamma,\sigma}N^{-1/3},$$

where we take $\alpha = 1+r$, $\beta = \sigma\gamma$, $K = 2r = \mathcal{O}(1/N)$, $x = (\gamma-\zeta)/36$, $c_4 = 1$ and $c_5 = 3/4$. We combine the above and (4.7) to arrive at the desired conclusion of (c).

For (d), we observe that the jump rates of $X_{\bullet 1} = X_{11} + X_{01}$ satisfy

$$\begin{aligned} r_{\bullet 1}^+ &= NX_{11}[(1+\sigma+r)X_{10} + (1+\sigma(1+\gamma)+2r)X_{00}] \\ &\quad + NX_{01}[(1+\sigma(1-\gamma)+2r)X_{10} + (1+\sigma+r)X_{00}] \\ r_{\bullet 1}^- &= NX_{11}[(1-\sigma+r)X_{10} + (1-\sigma(1+\gamma)+r)X_{00}] \\ &\quad + NX_{01}[(1-\sigma(1-\gamma)+r)X_{10} + (1-\sigma+r)X_{00}], \end{aligned}$$

where we drop the terms involving $X_{11}X_{01}$ in $r_{01}^{\pm}$ and $r_{11}^{\pm}$, which correspond to type 11 individuals replaced by type 01 individuals or *vice versa*. Therefore $X_{\bullet 1}$ dominates $1-\eta$ where we define η to be a jump process with initial condition $\eta(T_{01;\epsilon}) = 1 - X_{\bullet 1}(T_{01;\epsilon})$ and jump rates of

$$\begin{aligned} r_\eta^+ &= N(1-\sigma(1-\gamma)+r)\eta(1-\eta), \\ r_\eta^- &= N(1+\sigma(1-\gamma)+r)\eta(1-\eta). \end{aligned}$$

Since $\eta(T_{Z_{01};1-\epsilon}) \le 1 - X_{01}(T_{Z_{01};1-\epsilon}) \le \epsilon$ on $E_0 \cap G_2 \cap G_1 \cap E_1$, by Lemma 5.3,

$$\mathbb{P}\left(\{\eta(t) \ge \sqrt{\epsilon} \text{ for some } t \ge T_{Z_{01};1-\epsilon}\} \cap E_0 \cap G_2 \cap G_1 \cap E_1\right) \le C_{\gamma,\sigma} N^{-1/2},$$

where we take $\alpha = 1 + r$, $\beta = \sigma(1-\gamma)$, $K = 0$, $x = (\gamma - \zeta)/144$, $c_4 = 1$ and $c_5 = 1/2$. This implies the desired conclusion of (d).

Let $\tilde{\eta}$ be a time change of η by $1-\eta$, then $2N\tilde{\eta}$ is a branching process and the clock for $\tilde{\eta}$ runs at the rate of at most $1/(1-\tilde{\eta}) < 1.02$ times that of η on $\{\tilde{\eta}(t) < \sqrt{\epsilon}$ for all $t \ge T_{Z_{01};1-\epsilon}\} \cap E_0 \cap G_2 \cap G_1 \cap E_1$. By Lemma 6.1(b),

$$\mathbb{P}(\{\tilde{\eta}\,(T_{Z_{01};1-\epsilon} + 0.99 t_{late}) > 0\} \cap E_0 \cap G_2 \cap G_1 \cap E_1) \le CN\epsilon e^{-\log(2N)}.$$

Hence $\mathbb{P}\left(\{\eta\,(T_{Z_{01};1-\epsilon} + t_{late}) > 0\} \cap E_0 \cap G_2 \cap G_1 \cap E_1\right) \le C\epsilon$, which implies (e) since $T_{Z_{01};1-\epsilon} + t_{late} = T_\infty$. □

Proof of Proposition 3.1(g-h). In (g), we couple X_{11} and Z_{11} and show that they are likely to agree before the establishment time of X_{11} and T_∞. In (h), we time change Z_{11} to obtain a branching process, which is likely to exit from the interval $(0, \theta_{11})$ before too long. Let

$$S_{X,Z,far} = \inf\{t \ge T_{01;\epsilon} : |X_{01}(t) - Z_{01}(t)| \vee X_{00}(t) > \epsilon^{1/3}\}.$$

By Proposition 3.1(c,d), there exists $\delta > 0$ such that

$$\begin{aligned} &\mathbb{P}(\{S_{X,Z,far} \le T_{11;\theta_{11}}\} \cap G_1 \cap E_1) \\ &\quad \le \mathbb{P}((G_2^c \cup (G_3^c \cap E_0 \cap G_2)) \cap G_1 \cap E_1) \le C_{\gamma,\sigma} N^{-\delta}, \end{aligned} \tag{4.8}$$

where we have used that on $E_0 \cap G_2$, $S_{X,Z,far} \ge T_{Z_{01};1-\epsilon}$ and on G_3, $X_{01}(t) > 1 - \epsilon^{1/2} - X_{11}(t) > 1 - \epsilon^{1/2} - \theta_{11}$ and $X_{00}(t) \le 1 - X_{01}(t) - X_{11}(t) < \epsilon^{1/2}$ for $t \ge T_{Z_{01};1-\epsilon}$ ($Z_{01} = 1$ for such t). Notice that on $G_1 \cap E_1$, $X_{11}(t) = 0 = Z_{11}(t)$ for all $t \le T_{01;\epsilon}$. For $t < S_{X_{11},Z_{11},diff} \wedge S_{X,Z,far} \wedge T_{11;\theta_{11}}$, since $\sigma \in [0,1]$ and $r = \mathcal{O}(1/N) \le \theta_{11}/10 \le \epsilon^{1/3}/100$ for large N, we have

$$\begin{aligned} |r_{Z,11}^+ - r_{11}^+| &\le N\theta_{11}[(\sigma - r)2\epsilon^{1/3} + \theta_{11}] + 2rN(3\epsilon^{1/3} + \theta_{11}) \\ &\le N\theta_{11}[2\epsilon^{1/3} + \theta_{11}] + \frac{1}{5}\theta_{11} N(3\epsilon^{1/3} + \theta_{11}) \\ &= N\theta_{11}[2\epsilon^{1/3} + \theta_{11} + \frac{1}{5}(3\epsilon^{1/3} + \theta_{11})] \le 4N\theta_{11}\epsilon^{1/3}, \end{aligned}$$

and similarly, $|r_{Z,11}^- - r_{11}^-| \le 4N\theta_{11}\epsilon^{1/3}$. Thus the absolute difference between X_{11} and Z_{11} is bounded above by a Poisson process η that has initial value $\eta(T_{01;\epsilon}) = 0$ and jumps at rate $8N\theta_{11}\epsilon^{1/3}$. If $t_{mid} + t_{late} \le \epsilon^{-1/12}$, which is satisfied by our choice of $t_{mid} + t_{late} = \mathcal{O}(\log N)$, then

$$\mathbb{P}(\eta(s) > 0 \text{ for some } s \in [T_{01;\epsilon}, T_{01;\epsilon} + t_{mid} + t_{late}]) \le 1 - e^{-8N\theta_{11}\epsilon^{1/3-1/12}} < \epsilon^{1/6}.$$

On $E_2 \cap E_1$, the process η remaining 0 during $[T_{01;\epsilon}, T_{01;\epsilon} + t_{mid} + t_{late}]$ implies that X_{11} and Z_{11} are equal as long as we have not reached $S_{X,Z,far} \wedge T_{11;\theta_{11}}$, hence

$$\mathbb{P}(\{S_{X_{11},Z_{11},diff} \le T_\infty \wedge S_{X,Z,far} \wedge T_{11;\theta_{11}}\} \cap G_1 \cap E_1) \le \epsilon^{1/6}.$$

We combine (4.8) and the above estimate to obtain

$$\mathbb{P}(\{S_{X_{11},Z_{11},diff} \leq T_\infty \wedge T_{11;\theta_{11}}\} \cap G_1 \cap E_1) \leq \epsilon^{1/6} + C_{\gamma,\sigma} N^{-\delta},$$

which implies (g).

Let $F_4 = \{T_{Z_{11};\{0,\theta_{11}\}} \geq T_{Z_{01};1-\epsilon}\}$. Starting from $T_{Z_{01};1-\epsilon}$, Z_{11} is a time-changed branching process. We perform a time change of $1 - Z_{11}$ (from time $T_{Z_{01};1-\epsilon}$ onwards) to obtain a branching process $\tilde{Z}_{11}$, then the clock for $\tilde{Z}_{11}$ runs faster than that of Z_{11} (at a rate of at most $1/(1-\theta_{11})$ times before $\tilde{Z}_{11}$ reaches θ_{11}). From time $T_{Z_{01};1-\epsilon}$ onwards, 0 and θ_{11} are absorption points for $Z_{11}(\cdot \wedge T_{Z_{11};\theta_{11}})$. We use Lemma 6.1(d) below to deduce that

$$\begin{aligned}
&\mathbb{P}\left(\{Z_{11}(T_\infty \wedge T_{Z_{11};\theta_{11}}) \in (0,\theta_{11})\} \cap F_4 \cap G_1 \cap E_1\right) \\
&\quad\leq \mathbb{P}\left(\{\tilde{Z}_{11}(s) \in (0,\theta_{11}) \text{ for all } s \leq (1-\theta_{11})T_\infty\} \cap F_4 \cap G_1 \cap E_1\right) \\
&\quad\leq (2N\theta_{11})^2 C_{\gamma,\sigma} \exp(-0.99\sigma\gamma(T_\infty - T_{Z_{01};1-\epsilon})), \\
&\quad\leq C_{\gamma,\sigma}(\log^2 N)\exp(-0.99\sigma\gamma t_{late}) \leq C_{\gamma,\sigma} N^{-1/2}.
\end{aligned}$$

Therefore

$$\mathbb{P}\left(\{Z_{11}(T_\infty) \in (0,\theta_{11}),\ T_{Z_{11};\theta_{11}} > T_\infty\} \cap F_4 \cap G_1 \cap E_1\right) \leq C_{\gamma,\sigma} N^{-1/2}.$$

On $\{S_{X_{11},Z_{11},diff} > T_\infty \wedge T_{11;\theta_{11}}\}$, X_{11} and Z_{11} agree up to $T_\infty \wedge T_{11;\theta_{11}}$. Therefore

$$\begin{aligned}
\mathbb{P}\big(\{S_{X_{11},Z_{11},diff} > T_\infty, T_{11;\theta_{11}} > T_\infty, X_{11}(T_\infty) = Z_{11}(T_\infty) \in (0,\theta_{11}),& \\
T_{11;\{0,\theta_{11}\}} \geq T_{Z_{01};1-\epsilon}\} \cap G_1 \cap E_1\big) &\leq C_{\gamma,\sigma} N^{-1/2}.
\end{aligned}$$

We can drop the condition $T_{11;\{0,\theta_{11}\}} \geq T_{Z_{01};1-\epsilon}$, since on $\{S_{X_{11},Z_{11},diff} > T_\infty, T_{11;\theta_{11}} > T_\infty, T_{11;\{0,\theta_{11}\}} < T_{Z_{01};1-\epsilon}\}$, we have $X_{11}(T_\infty) = Z_{11}(T_\infty) = 0$. Hence

$$\mathbb{P}\left(\{T_{11;\theta_{11}} > T_\infty, X_{11}(T_\infty) = Z_{11}(T_\infty) \in (0,\theta_{11})\} \cap G_5 \cap G_1 \cap E_1\right) \leq C_{\gamma,\sigma} N^{-1/2},$$

which implies the desired result in (h). □

5 Supporting Lemmas

In this section, we establish Lemmas 5.1 to 5.3, one each for the early, middle, and late phases. They are used for the proof of Proposition 3.1 in §4. Lemma 5.1 deals with the early phase and approximates a 1-dimensional jump process undergoing selection (see (5.1) for a precise definition) by a deterministic function, where the error bound depends only on the initial condition of the process, as long as the process is stopped before it reaches $\mathcal{O}(1)$. Lemma 5.2 deals with the middle phase and uses the logistic growth as an approximation. The main difference between the early phase and the middle phase is the error bound. In Lemma 5.2, the error bound depends on both the initial and terminal conditions of the process. Lemma 5.3 deals with the late phase, for which we only need to show that the proportion of advantaged types does not stray too far away from 1 (or 0 for proportion of disadvantageous types) once it gets close to 1 (or 0).

Lemma 5.1. *Let $\alpha \geq 1$, $\beta \in (0,1)$, $\delta_0 \in [0,1/2]$ and $x \in (0,1]$ be constants. Let ξ be a jump process with initial value $\xi(0) = (2N)^{-x} \geq (2N)^{-1}$, jump size $1/(2N)$, and jump rates*

$$r^+ = N\xi[(\alpha+\beta)(1-\xi) - \delta_0],\ r^- = N\xi[(\alpha-\beta)(1-\xi) + \delta_0]. \tag{5.1}$$

Suppose Y is a deterministic process that satisfies

$$Y(t) = (2N)^{-x} + \int_0^t Y(s)(\beta(1-Y(s)) - \delta_0)\ ds.$$

If $\delta_1 \in (0,1)$ is a constant and $u_0 = \inf\{t : Y(t) = \delta_1\} \le (\log 2)/(3\beta\delta_1 + \delta_0)$, then there exists $\delta_2 \in (0,(1-x)/4]$ such that

$$\mathbb{P}\left(|\xi(s) - Y(s)| > 4N^{-\delta_2}Y(s) \text{ for some } s \le u_0\right) \le C_{\alpha,\beta}N^{-\delta_2}.$$

Moreover, if $\check{\xi}$ and $\hat{\xi}$ are jump processes such that $\hat{\xi} \ge \xi \ge \check{\xi}$ before a stopping time T, then $\mathbb{P}\left(\hat{\xi}(s) < (1-4N^{-\delta_2})Y(s) \text{ for some } s \le u_0 \wedge T\right) \le C_{\alpha,\beta}N^{-\delta_2}$ and $\mathbb{P}\left(\check{\xi}(s) > (1+4N^{-\delta_2})Y(s) \text{ for some } s \le\right.$ $C_{\alpha,\beta}N^{-\delta_2}$.

Proof. We can write

$$\begin{aligned} d\xi &= dM_\xi + \xi(\beta(1-\xi) - \delta_0)\ dt, \\ d\langle M_\xi\rangle &= \frac{\alpha}{2N}\xi(1-\xi)\ dt, \end{aligned}$$

and consequently,

$$\begin{aligned} d(e^{-\beta t}\xi(t)) &= d\tilde{M}_\xi(t) - e^{-\beta t}(\beta\xi(t)^2 + \delta_0\xi(t))\ dt \\ d\langle\tilde{M}_\xi\rangle(t) &= \frac{\alpha}{2N}e^{-2\beta t}\xi(t)(1-\xi(t))\ dt. \end{aligned} \tag{5.2}$$

We define $\tau = \inf\{t \le u_0 : \xi(t) \ge 2\delta_1\}$, and take expectation on both sides of (5.2) to obtain

$$E[e^{-\beta(t\wedge\tau)}\xi(t\wedge\tau)] = (2N)^{-x} - E\left[\int_0^{t\wedge\tau} e^{-\beta s}(\beta\xi(s)^2 + \delta_0\xi(s))\ ds\right] \le (2N)^{-x}.$$

As in the steps leading to (3.5), we use Jensen's and Burkholder's inequalities to obtain

$$\begin{aligned} E\left[\sup_{s\le t\wedge\tau}|\tilde{M}_\xi(s)|\right] &\le \frac{C}{N} + \frac{C_\alpha}{N^{1/2}}\left(E\left[\int_0^t e^{-2\beta s}\xi(s)\mathbf{1}_{\{s\le\tau\}}\ ds\right]\right)^{1/2} \\ &\le \frac{C}{N} + \frac{C_\alpha}{N^{1/2}}\left(\int_0^t e^{-\beta s}(2N)^{-x}\ ds\right)^{1/2} \le C_{\alpha,\beta}N^{-(1+x)/2}. \end{aligned} \tag{5.3}$$

Since $de^{-\beta t}Y(t) = -e^{-\beta t}(\beta Y(t)^2 + \delta_0 Y(t))\ dt$, we use (5.3) in (5.2) to obtain

$$\begin{aligned} &E\left[\sup_{s\le t\wedge\tau} e^{-\beta s}|\xi(s) - Y(s)|\right] \\ &\le\ C_{\alpha,\beta}N^{-(1+x)/2} + E\left[\int_0^{t\wedge\tau} e^{-\beta s}(\beta|\xi(s)^2 - Y(s)^2| + \delta_0|\xi(s) - Y(s)|)\ ds\right] \\ &\le\ C_{\alpha,\beta}N^{-(1+x)/2} + E\left[\int_0^t (3\beta\delta_1 + \delta_0)e^{-\beta s}|\xi(s) - Y(s)|\mathbf{1}_{\{s\le\tau\}}\ ds\right] \\ &\le\ C_{\alpha,\beta}N^{-(1+x)/2} + \int_0^t (3\beta\delta_1 + \delta_0)E\left[\sup_{s\le s'\wedge\tau} e^{-\beta s}|\xi(s) - Y(s)|\right]\ ds'. \end{aligned}$$

Gronwall's inequality implies

$$E\left[\sup_{s\le t\wedge\tau} e^{-\beta s}|\xi(s) - Y(s)|\right] \le C_{\alpha,\beta}N^{-(1+x)/2}e^{(3\beta\delta_1+\delta_0)t} \le C_{\alpha,\beta}N^{-(1+x)/2},$$

since $\tau \le u_0 \le (\log 2)/(3\beta\delta_1+\delta_0)$. Let $\delta_2 \in (0,(1-x)/4]$, then

$$\mathbb{P}\left(|\xi(s)-Y(s)| \ge N^{-\delta_2-x}e^{\beta s} \text{ for some } s \le u_0 \wedge \tau\right) \le C_{\alpha,\beta}N^{-\delta_2}.$$

We observe that for $s \le u_0$, $(2N)^{-x}e^{(\beta-\beta\delta_1-\delta_0)s} \le Y(s)$, hence $N^{-x}e^{\beta s}/Y(s) \le 2^x e^{(\beta\delta_1+\delta_0)s} \le 2^x e^{(\beta\delta_1+\delta_0)(\log 2)/(3\beta\delta_1+\delta_0)} \le 4$, i.e. $N^{-x}e^{\beta s} \le 4Y(s)$. Hence

$$\mathbb{P}\left(|\xi(s)-Y(s)| \ge 4N^{-\delta_2}Y(s) \text{ for some } s \le u_0 \wedge \tau\right) \le C_{\alpha,\beta}N^{-\delta_2}.$$

We can drop τ in the event above, since $|\xi(\tau)-Y(\tau)| \ge Y(\tau)$. The conclusion follows. □

Lemma 5.2. *Let $\beta,\epsilon_0,\epsilon_1 \in (0,1)$ and $a_0,a_1 > 0$ be constants. Suppose Y is a deterministic process defined from a stopping time u_1 onwards that has initial condition $Y(u_1) = b_0 \ge a_0(2N)^{-\epsilon_0}$ and satisfies*

$$Y(t) = b_0 + \int_{u_1}^t \beta Y(s)(1-Y(s))\,ds.$$

Let $u_2 = u_1 + \frac{1}{\beta}\log\frac{1-b_0}{b_0}\frac{1-b_1}{b_1}$ such that $Y(u_2) = 1-b_1 \le 1-a_1(2N)^{-\epsilon_1}$. Suppose T is a stopping time and ξ is a jump process that takes values in $[0,1]$, has jump size $1/(2N)$ and satisfies

$$\begin{aligned}\xi(t\wedge T) &= \xi(u_1) + M(t\wedge T) + \int_{u_1}^{t\wedge T}\xi(s)[\beta(1-\xi(s))+\epsilon_2(s)] + \epsilon_3(s)\,ds\\ \langle M\rangle(t\wedge T) &= \frac{1+r}{2N}\int_{u_1}^{t\wedge T}\xi(s)(1-\xi(s)) + \epsilon_4(s)\,ds,\end{aligned}$$

where $|\epsilon_2(t)|,|\epsilon_3(t)| \le (2N)^{-\delta_1}$, $\epsilon_4(t) \le 1$ for $t \le T$, and M is a jump martingale with jump size $1/(2N)$. Furthermore, suppose on a set $D_1 \in \mathcal{F}(u_1)$, we have $|\xi(u_1)-b_0| \le (2N)^{-\delta_2}$. We define $D_2 = \{T \ge u_1\}$ and δ_3 to be a constant $\le ((\delta_1\wedge\delta_2\wedge\frac{1}{2})-\epsilon_0-\epsilon_1)/3$, then we have

$$\mathbb{P}\left(\left\{\sup_{s\in[u_1,u_2\wedge T]}|\xi(s,\omega)-Y(s,\omega)| > (2N)^{-\delta_3}\right\}\cap D_1\cap D_2\right) \le (2N)^{-\delta_3}.$$

Proof. Let $D = D_1 \cap D_2$. Notice that $D \in \mathcal{F}(u_1)$. Since

$$\begin{aligned}&|\xi(t)[\beta(1-\xi(t))+\epsilon_2(t)] - \beta Y(t)(1-Y(t))|\mathbf{1}_{\{t\le T\}}\\ &\quad \le (2N)^{-\delta_1} + \beta|\xi(t)-Y(t)||1-\xi(t)-Y(t)|\mathbf{1}_{\{t\le T\}}\\ &\quad \le (2N)^{-\delta_1} + \beta|\xi(t)-Y(t)|\mathbf{1}_{\{t\le T\}},\end{aligned}$$

we have

$$\begin{aligned}&|\xi((u_1+t)\wedge T) - Y((u_1+t)\wedge T)|\mathbf{1}_D \le |\xi(u_1)-Y(u_1)|\\ &\qquad + |M((u_1+t)\wedge T)\mathbf{1}_D| + \int_{u_1}^{(u_1+t)\wedge T}[(2N)^{-\delta_1} + \beta|\xi(s)-Y(s)|]\mathbf{1}_D\,ds.\end{aligned}$$

By Jensen's and Burkholder's inequalities (C changes line to line below),

$$E\left[\sup_{u_1\le s\le u_1+t}|M(s\wedge T)\mathbf{1}_D|\right] \le \frac{C}{N} + C\sqrt{\frac{t}{N}} \le C\sqrt{\frac{t}{N}},$$

therefore

$$E\left[\sup_{u_1\le s\le u_1+t}|\xi(s\wedge T)-Y(s\wedge T)|\mathbf{1}_D\right]\le C\sqrt{\frac{t}{N}}+E\left[|\xi(u_1)-Y(u_1)|\mathbf{1}_D\right]$$
$$+2(2N)^{-\delta_1}t+\int_{u_1}^{u_1+t}\beta E[|\xi(s)-Y(s)|\mathbf{1}_{\{s\le T\}}\mathbf{1}_D]\,ds.$$

Since $E[|\xi(s)-Y(s)|\mathbf{1}_{\{s\le T\}}\mathbf{1}_D]\le E[|\xi(s\wedge T)-Y(s\wedge T)|\mathbf{1}_D]$, and $|\xi(u_1)-Y(u_1)|\mathbf{1}_D\le(2N)^{-\delta_2}$, we have

$$E\left[\sup_{u_1\le s\le u_1+t}|\xi(s\wedge T)-Y(s\wedge T)|\mathbf{1}_D\right]\le C\left(\sqrt{\frac{t}{N}}+\frac{1}{(2N)^{\delta_2}}+\frac{t}{(2N)^{\delta_1}}\right)e^{\beta t}$$

by Gronwall's inequality. We observe that $u_2-u_1\le\frac{1}{\beta}\log\frac{1}{b_0b_1}\le\frac{1}{\beta}[(\epsilon_0+\epsilon_1)\log(2N)-\log(a_0a_1)]$, therefore the estimate above implies

$$E\left[\sup_{u_1\le s\le u_2}|\xi(s\wedge T)-Y(s\wedge T)|\mathbf{1}_D\right]\le\frac{C(2N)^{\epsilon_0+\epsilon_1}\log N}{a_0a_1}(2N)^{-(\delta_1\wedge\delta_2\wedge\frac{1}{2})}.$$

Since $0<\delta_3\le((\delta_1\wedge\delta_2\wedge\frac{1}{2})-\epsilon_0-\epsilon_1)/3$, we have

$$E\left[\sup_{u_1\le s\le u_2}|\xi(s\wedge T)-Y(s\wedge T)|\mathbf{1}_D\right]\le(2N)^{-2\delta_3},$$

where $\frac{C\log N}{a_0a_1}<(2N)^{-\delta_3}$ for large N. This implies the desired conclusion. □

Lemma 5.3. *Let $\alpha\ge 1$, $\beta\in(0,1)$, $x\in(0,1]$, $c_4>0$ and $K\ge 0$ be constants. Let $\eta\le\hat\eta$ be jump processes where η has initial value $\eta(0)=1-c_4(2N)^{-x}$, jump size $1/(2N)$, jump rates*

$$r^+=N(\alpha+\beta)\eta(1-\eta),\ r^-=N(\alpha-\beta)\eta(1-\eta)+NK.$$

For $t\le c_4(2N)^{-x}/K$ (if $K=0$, then $t=\infty$), $c_5\in(0,1)$ and sufficiently large N, we have

$$\mathbb{P}\left(\inf_{s\le t}\hat\eta(s)>1-(2N)^{-c_5x}\right)\ge\mathbb{P}\left(\inf_{s\le t}\eta(s)>1-(2N)^{-c_5x}\right)\ge 1-C_{\alpha,\beta}N^{(c_5-1/2)x-1/2}.$$

Proof. We take $\xi=1-\eta$ and perform a time change of $1-\xi$ on ξ to obtain a process $\tilde\xi$ with jump rates

$$\tilde r^+=N(\alpha-\beta)\tilde\xi+NK/(1-\tilde\xi),\ \tilde r^-=N(\alpha+\beta)\tilde\xi.$$

Let $\tilde\xi_{up}$ be a jump process with initial condition $\tilde\xi_{up}(0)=\tilde\xi(0)=c_4(2N)^{-x}$, jump size $1/(2N)$ and jump rates

$$\tilde r^+_{up}=N(\alpha-\beta)\tilde\xi_{up}+2NK,\ \tilde r^-_{up}=N(\alpha+\beta)\tilde\xi_{up}.$$

Before the stopping time $\tau=\inf\{t\ge 0:\tilde\xi_{up}\ge 1/2\}$, $\tilde\xi_{up}$ dominates $\tilde\xi$. We can write

$$d\tilde\xi_{up}(t)=dM_{\tilde\xi_{up}}+(K-\beta\tilde\xi_{up})\,dt, d\langle M_{\tilde\xi_{up}}\rangle=\frac{1}{2N}(K+\alpha\tilde\xi_{up}(t))\,dt.$$

Hence $E[\tilde\xi_{up}(t)]=\frac{K}{\beta}+\left(c_4(2N)^{-x}-\frac{K}{\beta}\right)e^{-\beta t}$ and by Jensen's and Burkholder's inequalities,

$$\begin{aligned}E\left[\sup_{s\le 2t}M_{\tilde\xi_{up}}(s)\right]&\le\frac{C}{N}+\frac{C}{\sqrt N}\left(Kt+\alpha\int_0^{2t}E[\tilde\xi_{up}(t)]\,ds\right)^{1/2}\\&\le\frac{C_{\alpha,\beta}}{\sqrt N}\left(Kt+c_4(2N)^{-x}\right)^{1/2}\le C_{\alpha,\beta}N^{-(1+x)/2},\end{aligned}$$

if $Kt \le c_4(2N)^{-x}$, in which case

$$\begin{aligned}\mathbb{P}\left(\sup_{s\le 2t}\tilde\xi_{up}(s)\ge (2N)^{-c_5x}\right) &\le \mathbb{P}\left(\sup_{s\le 2t} M_{\tilde\xi_{up}}(s)\ge (2N)^{-c_5x}-c_4(2N)^{-x}-4Kt\right)\\ &\le \frac{C_{\alpha,\beta}N^{-(x+1)/2}}{(2N)^{-c_5x}-c_4(2N)^{-x}-4Kt}\\ &\le C_{\alpha,\beta}N^{(c_5-1/2)x-1/2}.\end{aligned}$$

On the set $\{\sup_{s\le 2t}\tilde\xi_{up}(s)\le (2N)^{-c_5x}\}$, $\tilde\xi_{up}$ certainly does not reach $1/2$ before time $2t$. Hence $\tilde\xi_{up}$ dominates $\tilde\xi$ before $2t$ for $\omega\in\{\sup_{s\le 2t}\tilde\xi_{up}(s)\le (2N)^{-c_5x}\}$, which implies $\mathbb{P}\left(\sup_{s\le 2t}\tilde\xi(s)<(2N)^{-c_5x}\right)\ge 1-C_{\alpha,\beta}N^{(c_5-1/2)x-1/2}$. Because $\tilde\xi$ is the process ξ after a time change of $1-\xi$, the clock for $\tilde\xi$ runs faster than that of ξ, but at most twice as fast before $\tilde\xi$ reaches $1/2$. Therefore the estimate above implies $\mathbb{P}\left(\sup_{s\le t}\xi(s)<(2N)^{-c_5x}\right)\ge \mathbb{P}\left(\sup_{s\le 2t}\tilde\xi(s)<(2N)^{-c_5x}\right)\ge 1-C_{\alpha,\beta}N^{(c_5-1/2)x-1/2}$. The conclusion follows. □

6 Appendix: A Result on Branching Processes

Lemma 6.1. *Let $\xi^{(k)}$ be a branching process with $\xi(0)=k$ and $u(s)=as^2+b$ be the probability generating function of the offspring distribution. Then*

$$G(s,t)=E(s^{\xi^{(k)}(t)})=\left(\frac{b(s-1)-(as-b)e^{-(a-b)t}}{a(s-1)-(as-b)e^{-(a-b)t}}\right)^k.$$

(a) If $k=1$ and $a>b$, then

1. $|\mathbb{P}(\xi^{(1)}(t)=0)-b/a|\le be^{-(a-b)t}/a$.
2. $\mathbb{P}(1\le \xi^{(1)}(t)\le K)\le C_{a,b}Ke^{-(a-b)t}$ *if* $K\le e^{(a-b)t}/6$.
3. $\mathbb{P}\left(\sup_{s\le t}\xi^{(1)}(s)\ge K\right)\le C_{a,b}e^{(a-b)t}/K$.

(b) If $a<b$, then $\mathbb{P}(\xi^{(k)}(t)>0)\le 1.2ke^{-(b-a)t}$.
(c) If $a>b$ and $k\in[1,K]$, then $\mathbb{P}(\xi^{(k)}(t)\in[1,K])\le kC_{a,b}Ke^{-(a-b)t}$.
(d) If $a>b$ and ξ is a branching process with an initial condition that has support on $[0,k]$, then $\mathbb{P}(\xi(t)\in[1,K])\le kC_{a,b}Ke^{-(a-b)t}$. Consequently,

$$\mathbb{P}(\xi(s)\in[1,K] \text{ for all } s\le t)\le kC_{a,b}Ke^{-(a-b)t}.$$

Proof. The formula for $G(s,t)$ comes from Chapter III.5 of Athreya & Ney (1972). From this formula, we deduce that

$$\mathbb{P}(\xi^{(k)}(t)=0)=G(0,t)=\left(\frac{b-be^{-(a-b)t}}{a-be^{-(a-b)t}}\right)^k. \tag{6.1}$$

For (a), we specialise to the case of $k=1$ and $a>b$. We write $\xi=\xi^{(1)}$, then

$$\left|\mathbb{P}(\xi(t)=0)-\frac{b}{a}\right|=\frac{(a-b)be^{-(a-b)t}}{a(a-be^{-(a-b)t})}\le \frac{b}{a}e^{-(a-b)t},$$

as required by (a.1). For $s\le 1$, we have

$$\begin{aligned}\mathbb{P}(1\le\xi(t)\le K)&\le s^{-K}\sum_{i=1}^{\infty}\mathbb{P}(\xi(t)=i)s^i=s^{-K}(G(s,t)-G(0,t))\\ &=\frac{(a-b)^2s}{(a-be^{-(a-b)t})(a(e^{(a-b)t}s^K(1-s)+s^{K+1})-bs^K)}.\end{aligned}$$

where $G(s,t) - G(0,t)$ can be computed from (6.1) using elementary algebra. The dominant term in the denominator of the above quantity is $e^{(a-b)t}s^K(1-s)$, which achieves the maximum

$$\frac{e^{(a-b)t}}{K+1}\left(1-\frac{1}{K+1}\right)^K = \frac{e^{(a-b)t}}{K}\left(1-\frac{1}{K+1}\right)^{K+1}$$

at $s = K/(K+1)$. For sufficiently large K, this is at least $e^{(a-b)t}/(3K)$. Therefore

$$\mathbb{P}(1 \le \xi(t) \le K) \le \frac{(a-b)^2\frac{K}{K+1}}{(a-be^{-(a-b)t})\left(a\frac{e^{(a-b)t}}{3K}-b\right)} \le C_{a,b}\left(a\frac{e^{(a-b)t}}{3K}-b\right)^{-1},$$

which implies the desired conclusion of (a.2), if $K \le e^{(a-b)t}/6$.

For (a.3), we observe that $M(t) = e^{-(a-b)t}\xi(t)$ is a martingale with maximum jump size 1 and quadratic variation $\langle M\rangle(t) = \int_0^t e^{-2(a-b)s}(a+b)\xi(s)\,ds$. Burkholder's inequality implies

$$\begin{aligned} E\left[\sup_{s\le t} M(s)\right] &\le C + C\int_0^t e^{-2(a-b)s}(a+b)E[\xi(s)]\,ds \\ &= C + C\int_0^t e^{-2(a-b)s}(a+b)e^{(a-b)s}\,ds \le C_{a,b}. \end{aligned}$$

Therefore $E\left[\sup_{s\le t}\xi(s)\right] \le C_{a,b}e^{(a-b)t}$, which implies (a.3).

For (b), we observe that

$$\mathbb{P}(\xi^{(k)}(t) = 0) = \left(1-\frac{b-a}{be^{(b-a)t}-a}\right)^k.$$

For sufficiently large t, we have

$$\frac{be^{(b-a)t}-a}{b-a} = \frac{e^{(b-a)t}-\frac{a}{b}}{1-\frac{a}{b}} \ge e^{(b-a)t} - \frac{a}{b} \ge \frac{1}{1.1}e^{(b-a)t},$$

therefore

$$\begin{aligned} \mathbb{P}(\xi^{(k)}(t)=0) &\ge \left[\left(1-\frac{b-a}{be^{(b-a)t}-a}\right)^{\frac{be^{(b-a)t}-a}{b-a}}\right]^{k(1.1)e^{-(b-a)t}} \ge e^{-1.2ke^{-(b-a)t}} \\ &\ge 1-1.2ke^{-(b-a)t}, \end{aligned}$$

if t is sufficiently large and $ke^{-(b-a)t}$ is sufficiently small.

For (c), we observe that $\xi^{(k)} = \xi_1^{(1)} + \xi_2^{(1)} + \ldots + \xi_k^{(1)}$, where $\xi_i^{(1)}, i = 1,\ldots,k$ are independent copies of $\xi^{(1)}$. Therefore

$$\mathbb{P}(\xi^{(k)}(t) \in [1,K]) \le \mathbb{P}(\xi_i^{(1)} \in [1,K] \text{ for some } i = 1,\ldots,k) \le kC_{a,b}Ke^{-(a-b)t}$$

by part (a.2) of this lemma. Part (d) is a direct consequence of part (c). □

References

[1] K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040]. MR-0373040

[2] N. H. Barton. Linkage and the limits to natural selection. *Genetics*, 140:821–841, 1995.

[3] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability*, 1:19–42, 1973. MR-0365692

[4] A. M. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. An approximate sampling formula under genetic hitchhiking. *Ann. Appl. Probab.*, 16(2):685–729, 2006. MR-2244430

[5] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. MR-0838085

[6] W. J. Ewens. *Mathematical Population Genetics, 2nd Edition*. Springer, 2004. MR-0554616

[7] R. A Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, 1930. MR-1785121

[8] J. H. Gillespie. Genetic drift in an infinite population: The pseudohitchiking model. *Genetics*, 155:909–919, 2000.

[9] J. H. Gillespie. Is the population size of a species relevant to its evolution? *Evolution*, 55:2161–2169, 2001.

[10] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. The Clarendon Press, Oxford University Press, New York, second edition, 1992. MR-1199812

[11] J. B. Haldane. The mathematical theory of natural and artificial selection. *Proc. Camb. Philos. Soc.*, 23:838:844, 1927.

[12] W. G. Hill and A. Robertson. The effect of linkage on limits to artificial selection. *Genetics Research*, 8:269–294, 1966.

[13] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland mathematics library, 1981. MR-1011252

[14] Samuel Karlin and Howard M. Taylor. *A second course in stochastic processes*. Academic Press Inc., New York, 1981. MR-0611513

[15] Y. Kim. Allele frequency distribution under recurrent selective sweeps. *Genetics*, 172:1967–1978, 2006.

[16] M. Kimura and T. Ohta. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763, 1969.

[17] G. A. McVean and Charlesworth B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and recombination. *Genetics*, 155:929–44, 2000.

[18] H. J. Muller. Some genetic aspects of sex. *American Naturalist*, 66:118–138, 1932.

[19] S. P. Otto and N. H. Barton. The evolution of recombination: removing the limits to natural selection. *Genetics*, 147:879–906, 1997.

[20] Feng Yu and A. M. Etheridge. The fixation probability of two competing beneficial mutations. *Theor. Popul. Biol.*, 78:36–45, 2010.