Branching with Selection and Mutation I: Mutant Fitness of Fréchet Type

Su-Chan Park[1] · **Joachim Krug[2]** · **Léo Touzo[3]** · **Peter Mörters[4]**

Abstract
We investigate two stochastic models of a growing population with discrete and non-overlapping generations, subject to selection and mutation. In our models each individual carries a fitness which determines its mean offspring number. Many of these offspring inherit their parent's fitness, but some are mutants and obtain a fitness randomly sampled, as in Kingman's house-of-cards model, from a distribution in the domain of attraction of the Fréchet distribution. We give a rigorous proof for the precise rate of superexponential growth of these stochastic processes and support the argument by a heuristic and numerical study of the mechanism underlying this growth. This study yields in particular that the empirical fitness distribution of one model in the long time limit displays periodic behaviour.

Keywords Branching process · Random environment · Population model · Selection · Mutation · Survival of the fittest · Fréchet distribution

---

Communicated by Hal Tasaki.

✉ Peter Mörters
  p.moerters@uni-koeln.de

  Su-Chan Park
  spark0@catholic.ac.kr

  Joachim Krug
  jkrug@uni-koeln.de

  Léo Touzo
  leo.touzo@phys.ens.fr

[1] Department of Physics, The Catholic University of Korea, 43 Jibong-ro, Bucheon-si, Gyeonggi-do 14662, Republic of Korea

[2] Institut für Biologische Physik, Universität zu Köln, Zülpicher Str. 77, 50937 Cologne, Germany

[3] Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, 75005 Paris, France

[4] Mathematisches Institut, Universität zu Köln, Weyertal 86–90, 50931 Cologne, Germany

1 Introduction

While the theory of branching processes is undoubtedly one of the best developed areas of probability theory, stochastic branching models that incorporate effects of selection and mutation have only recently become the subject of rigorous mathematical analysis. This is despite the unquestionable relevance of these effects to the evolution of populations in nature and in the laboratory [1–3].

By contrast, deterministic high density models of a population undergoing selection and mutation have been studied for quite some time [4]. The model most closely associated with our stochastic process is Kingman's model [5]. This is a dynamical system on the space of probability measures describing the fitness distribution of a population. The fitness distribution p_t at generation t of the population is replaced in the next generation by

$$p_{t+1}(dx) = (1 - \beta) \frac{x \, p_t(dx)}{\int y \, p_t(dy)} + \beta \mu(dx). \tag{1}$$

Here a proportion $1 - \beta$ of the new generation has been selected from the current generation proportionally to their fitness and a proportion β are mutants that get a new fitness, sampled independently of their past using a fixed mutant fitness distribution (MFD) μ. Note that this model is only well-defined if the mean fitness of the population remains finite and therefore requires moment bounds for the MFD. For certain MFDs with bounded support Kingman's model undergoes a condensation phase transition, which implies that a nonzero fraction of the total population attains the maximally possible fitness value when the mutation rate is low enough [6]. A rigorous analysis of the condensation transition can be found in [7], and variants of the model have been considered in [8–10].

Kingman's model is based on two main assumptions about the evolutionary process. First, the fitness of a mutant is assumed to be random and independent of the parental fitness, a setting often referred to as the "house-of-cards" model [4, 5, 11]. Second, each mutation gives rise to a new genetic type or allele, an assumption known as the infinite alleles model [12]. Park and Krug [13] studied a stochastic version of Kingman's model which incorporates these two features. The population is updated in discrete generations following asexual Wright–Fisher dynamics, which can be viewed as a branching process conditioned on a constant population size N, see [1]. For unbounded MFDs the mean population fitness grows without bound. Since the MFD is time-independent, the fraction of beneficial mutations declines indefinitely over time. As a consequence, for long times beneficial mutants emerge and evolve independently of each other, and the dynamics can be analyzed rather straightforwardly in terms of a record-like point process [13]. In particular, for a MFD with an exponential tail the mean population fitness increases logarithmically, which is consistent with the behaviour observed in Lenski's long-term evolution experiment with bacteria [2, 14]. A generalization of this model that includes the response of the immune system to a population of pathogens was considered in [15].

The decline of the supply of beneficial mutations distinguishes the fixed population version of Kingman's model from a related class of stochastic models where the selection coefficients of novel mutants, rather than their fitness, is drawn from a given, time-invariant distribution [1]. In these models the fitness distribution of the population converges to a fitness wave traveling at a constant speed, which is determined by the interference between competing mutant clones [16–21].

The branching process version of Kingman's model considered in this paper is, in a sense, intermediate between the deterministic model (1) and the stochastic finite population model of [13]. The dynamics of the branching process is stochastic, but because of the

unbounded growth of the population, competing clones can coexist at arbitrary long times and the population retains a nontrivial type structure. While our motivation here is primarily conceptual and mathematical, we note that the clonal composition of growing populations is a problem of considerable interest for the modeling of proliferating tumours [22–24]. In this context, Durrett et al. [22] studied a branching process with selection and mutation where, similar to the fitness wave models described in the previous paragraph, the selection coefficients of beneficial mutations are drawn from a fixed, continuous probability distribution with bounded or unbounded support.

The first papers studying branching process models of Kingman type that express the selective advantage of a fit individual in terms of its offspring distribution are [25], which deals with Weibull type MFDs and puts the focus on the condensation phenomenon in that model, and [26] which looks at the growth of the fittest family in the case of Gumbel type MFDs. Both papers are limited to bounded MFDs and implicitly rely on the analogy to Kingman's original model, though of course the methods of study are entirely different in a stochastic setting. The present paper initiates the study of Kingman type branching processes with selection and mutation for unbounded MFDs. We focus on the case of Fréchet type MFDs, where the mathematical challenge is linked to the fact that the analogous Kingman model is ill-defined [13].

The structure of the paper is as follows. In Sect. 2, we introduce the models and state the main theorem. Section 3 explains the heuristics behind the formal results and in Sect. 4 we present a rigorous proof of the theorem. Section 5 contains refined results for the empirical fitness distribution for one of our models. These results are not yet accessible by a complete rigorous mathematical analysis, so that we resort to a numerical and heuristic study and a rigorous analysis of an approximating deterministic system. In Sect. 6 we provide a short discussion that places our results into the context of previous work and points to directions for future research.

2 Models and Main Result

We study two models of a population evolving in discrete and non-overlapping generations. In both models all individuals are assigned a fitness value, which is a positive real number. As model parameters we fix a probability distribution μ on $(0, \infty)$ from which the random fitness values F are sampled, referred to in the following as the mutant fitness distribution (MFD), and a mutation probability $\beta \in (0, 1)$. As was explained in the Introduction, we assume an infinite alleles model with a house-of-cards fitness landscape.

In both models we start from generation $t = 0$ with a single individual[1] with fitness f. Each individual in the population in generation $t \geq 0$ produces a Poisson random number of offspring with mean given by its fitness. With probability $1 - \beta$ an offspring individual inherits its parent's fitness and is added to the population at generation $t + 1$. Otherwise, with probability β, it is a mutant. The two models differ in the fate of the mutants.

- *Fittest mutant model (FMM)* Every mutant is assigned a fitness sampled independently from μ. Only the fittest mutant (if there is one) is added to the population at generation $t + 1$. All other mutants die instantly.
- *Multiple mutant model (MMM)* Every mutant is assigned a fitness sampled independently from μ and is added to the population at generation $t + 1$.

---

[1] The generalization to multiple individuals is straightforward.

We write $X(t)$ for the number of individuals in generation t and study the growth of the population conditioned on the event of survival, i.e. when $X(t) \neq 0$ for all times t. It is easy to see that the population size of the MMM dominates the population size of the FMM at all times. Because the growth is determined by the fittest mutants we expect both models to grow at the same rate and to show this, it suffices to find an upper bound for the MMM and a matching lower bound for the FMM.

Naturally, the rate of growth depends on the MFD μ. If μ is an unbounded distribution in both models individuals of ever increasing fitness occur and hence the population will grow superexponentially fast. By contrast, if μ is bounded we can only have exponential growth. Indeed, if μ is continuous with essential supremum one, then for a closely related continuous time model of immortal individuals, it is shown in [25, Remark 1] that

$$\lim_{t \to \infty} \frac{\log X(t)}{t} = \lambda^*,$$

where $\lambda^* \in [1 - \beta, 1)$ is the unique solution of the equation

$$\beta \int \frac{\lambda^*}{\lambda^* - (1 - \beta)x} \mu(dx) = 1$$

if $\beta \int \frac{1}{1-x} \mu(dx) \geq 1$, and otherwise $\lambda^* := 1 - \beta$. Further details on the long term growth of the process in [25] depend on the classification of μ according to its membership in the max domain of attraction of an extremal distribution. This also applies to other model variants and unbounded MFDs. By the celebrated Fisher-Tippett theorem there are three such universality classes, see for example [27, Proposition 0.3]. These are

- the *Weibull class*, which roughly occurs if μ is bounded with mass decaying slowly near the essential supremum,
- the *Gumbel class*, which roughly occurs if the mass of μ is decaying quickly near the essential supremum, which may be finite or infinite,
- the *Fréchet class*, which roughly occurs if μ is unbounded with mass decaying slowly near infinity.

Extreme value theory plays an important role in the interpretation of experimentally determined effect size distributions of beneficial mutations, and representatives of all three universality classes have been identified empirically [28–31].

In the present paper, we are mainly interested in the asymptotic behaviour of the population size $X(t)$ in the case that μ belongs to the Fréchet class (or, in short, is of *Fréchet type*). Precisely, this means that the tail function

$$G(x) := \mu((x, \infty)) = \mathbb{P}(F > x)$$

is regularly varying with index $-\alpha$ for some $\alpha > 0$. In other words, there exists a function $\ell \colon (0, \infty) \to \mathbb{R}$ which is slowly varying at infinity such that $G(x) = x^{-\alpha} \ell(x)$. MFDs of Fréchet type have been found in several experimental studies [32–35], and appear to be typical for populations subjected to strong selection pressures, such as bacteria or viruses exposed to drugs.

As in this case μ is an unbounded distribution, the process $(X(t) \colon t \geq 0)$ will grow superexponentially fast on survival and therefore our discussion will focus on the limiting quantity

$$\nu = \lim_{t \to \infty} \frac{\log \log X(t)}{t}.$$

Our main result is stated in the following theorem.

Theorem 1 *Given $\alpha > 0$, let $T \in \mathbb{N}$ be the unique number such that*

$$\frac{(T-1)^T}{T^{T-1}} < \alpha \leq \frac{T^{T+1}}{(T+1)^T}$$

and define (see Lemma 4 for another equivalent definition)

$$\nu(\alpha) := \frac{1}{T} \log \frac{T}{\alpha}. \tag{2}$$

Let $(X(t))_{t \geq 0}$ be the size of the population in either the FMM or the MMM. Then, almost surely on survival,

$$\lim_{t \to \infty} \frac{\log \log X(t)}{t} = \nu(\alpha).$$

We would like to emphasize that although the survival probability depends not only on the initial condition but also on the model (FMM or MMM), the almost sure convergence on survival in Theorem 1 holds irrespective of the actual value of the survival probability as long as it is nonzero. Before presenting the proof of the theorem in Sect. 4, in the next section we motivate the expression (2).

3 Motivation of the Main Result

Here we explain the statement of Theorem 1 by a heuristic analysis of the FMM. For convenience we take the MFD μ to be of Pareto form, $G(x) = x^{-\alpha}$ for $x \geq 1$ and $G(x) = 1$ for $x < 1$. Moreover, throughout this section we assume that the initial fitness f is so large that the fluctuations induced by Poisson sampling are negligible at all times, which implies that both the total population size and the sizes of subpopulations of mutants are well approximated by their expectations. Denoting the fitness of the mutant that is added to the population at generation t by W_t, we can then write

$$X(t) \approx (1-\beta)^t f^t + \sum_{i=1}^{t} (1-\beta)^{t-i} W_i^{t-i}, \tag{3}$$

where the factors $1 - \beta$ account for the fact that (apart from the added mutant) only the unmutated fraction of the population survives to the next generation. For the same reason the total number N_t of mutants produced in generation t (including the ones that die immediately) is approximately[2]

$$N_t \approx \frac{\beta}{1-\beta} X(t).$$

Since the probability that the largest fitness W_t among N_t independent and identically distributed random variables with common distribution G is smaller than x is $(1 - x^{-\alpha})^{N_t}$, the random variable W_t can be sampled as

$$W_t = \left(1 - Z_t^{1/N_t}\right)^{-1/\alpha} \approx X(t)^{1/\alpha} Y_t, \quad Y_t := \left(\frac{1-\beta}{\beta} \log \frac{1}{Z_t}\right)^{-1/\alpha},$$

---

[2] In the FMM, given total population size $X(t-1)$ and population's mean fitness $\bar{F}_{t-1}$ at generation $t-1$, we can approximate $N_t \approx \beta X(t-1)\bar{F}_{t-1}$ and $X(t) \approx (1-\beta)X(t-1)\bar{F}_{t-1}$. Therefore, to associate N_t with $X(t)$ correctly, we need $1 - \beta$ in the denominator.

where Z_t is uniformly distributed in the interval $(0, 1)$ and we have approximated $Z_t^{1/N_t} \approx 1 + (1/N_t) \log Z_t$. Note that Y_t does not depend on $X(t)$.

To proceed, we define ω_t as

$$\omega_t := \frac{\log X(t)}{\log f},$$

which implies that $X(t) = f^{\omega_t}$ and $W_t \approx Y_t f^{\omega_t/\alpha}$. Inserting these relations into (3) we obtain

$$f^{\omega_t} \approx (1-\beta)^t f^t + \sum_{i=1}^{t} (1-\beta)^{t-i} Y_i^{t-i} f^{(t-i)\omega_i/\alpha}. \tag{4}$$

In the limit $f \to \infty$ with t fixed, the sum on the right hand side is dominated by the term with the largest exponent of f. Correspondingly, the ω_t for large but finite f can be well approximated by the solution χ_t of the recursion relation

$$\chi_t = \max\left\{ t, \frac{t-1}{\alpha}\chi_1, \frac{t-2}{\alpha}\chi_2, \ldots, \frac{t-k}{\alpha}\chi_k, \ldots, \frac{1}{\alpha}\chi_{t-1} \right\} \tag{5}$$

with $\chi_1 = 1$. We now argue that the χ_t grow at least exponentially. Since for any $t_0 \geq 1$ and any positive integer m

$$\chi_{t_0+m} \geq \frac{m}{\alpha}\chi_{t_0},$$

we have, for any $n \geq 1$

$$\chi_{t_0+nm} \geq \left(\frac{m}{\alpha}\right)^n \chi_{t_0}.$$

Correspondingly

$$\lim_{t\to\infty} \frac{\log \chi_t}{t} = \lim_{n\to\infty} \frac{\log \chi_{t_0+nm}}{nm} \geq \frac{1}{m}\log\frac{m}{\alpha}, \tag{6}$$

where we have assumed that the limit is well-defined. Since (6) is valid for any integer $m \geq 1$, an optimal lower-bound can be found by maximizing the right hand side. As shown by Lemma 4 in Sect. 4, the maximizer over the positive integers is precisely the function $\nu(\alpha)$ in Theorem 1. As the population size depends exponentially on ω_t or χ_t, the heuristic argument makes it plausible that $\nu(\alpha)$ is a lower bound on the double-exponential growth rate of $X(t)$. Remarkably, Theorem 1 states that the bound is tight, and moreover applies also to the MMM. Informally this implies that the population at time t is dominated by the fittest mutant that was generated at time $t - T$. As a consequence the empirical fitness distribution changes periodically with period T (see Sect. 5 for further discussion).

In Fig. 1, we depict $\nu(\alpha)$ together with the numerical solution[3] of the recursion relation (5). The fact that $\nu(\alpha)$ is the exact exponential growth rate of χ_t is proven rigorously in Lemma 6 in Sect. 4. In the inset of Fig. 1, we compare (2) to an approximation obtained by treating m in (6) as a continuous variable. This yields

$$\max\left\{ \frac{1}{x}\log\frac{x}{\alpha} : x \geq 1 \right\} = \begin{cases} 1/(e\alpha), & \alpha e \geq 1, \\ -\log\alpha, & \alpha e < 1. \end{cases} \tag{7}$$

---

[3] The direct numerical estimate of ν by extrapolating $\log \chi_t/t$ is hampered by the fact that $\lim_{t\to\infty} \chi_t e^{-\nu t}$ does not exist in general (see Sect. 5). The method used to obtain the data in Fig. 1 is explained at the end of Sect. 5.1.

Fig. 1 Plots of ν vs α. Solid line depicts (2) and open circles present numerical solutions of the recursion relation (5). Inset: Plot of $(\nu e \alpha)^{-1} - 1$ vs. α with ν in (2). The error of the approximation (7) is small and vanishes when $\alpha \leq 1/e$
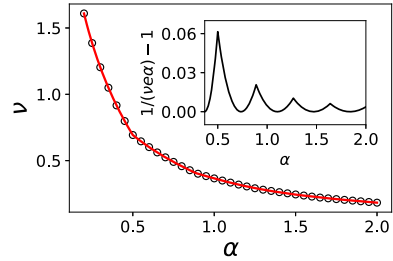
Although (7) is not exact, the relative error is less than 7% in all cases.

For $\alpha e < 1$ the expressions (2) and (7) actually coincide. In this regime of extremely heavy-tailed MFDs (more precisely, in the case of $\alpha \leq 0.5$ with $T = 1$; see Theorem 1) selection becomes irrelevant, in the sense that the double-exponential growth rate $\nu(\alpha) = \log(1/\alpha)$ persists in the extreme case $\beta \to 1$ of the MMM, where all individuals are replaced by mutants in each generation and the process becomes a classical Galton–Watson process albeit with infinite mean, cf. [36]. In the case of the FMM, the extreme case would stop the population from growing but the fitness W_t of the single individual present approximately satisfies $W_{t+1} \approx W_t^{1/\alpha}$, which gives

$$\lim_{t \to \infty} \frac{\log \log W_t}{t} = \log(1/\alpha).$$

4 Proof of Theorem 1

4.1 Preparation for the Proof

In this subsection we collect some tools that will be used in the proofs of the lower and upper bounds in the estimate leading to Theorem 1. The lower bound will be verified in Sect. 4.2 and the upper bound in Sect. 4.3.

For $t \in \mathbb{N}_0$ let W_t be the fitness of the fittest of the mutants in generation t and $W_t = 0$ if there are no mutants in generation t. Our first observation is that under the weak assumption $G(x) > 0$ for all large x (which always holds if μ is of Fréchet type) either the sequence (W_t) is unbounded or the branching process dies out in finite time. Heuristically speaking, on survival the accumulated number of mutants is unbounded almost surely, which naturally entails unbounded largest fitness.

Lemma 2 *Almost surely on survival the sequence (W_t) is unbounded.*

Proof We first show that the branching process can be coupled to a sequence $(\xi_1, \ldots, \xi_t)$ of independent Bernoulli variables with success parameter β and an independent sequence $(F_1, \ldots, F_t)$ of independent fitnesses with distribution μ such that *on survival up to generation t* we have, for all $1 \leq i \leq t$,

- $\xi_i = 1$ if there is at least one mutant in generation i, and
- $W_i \geq F_i \xi_i$.

Indeed, once the random variables $(F_1, \ldots, F_t)$ and $(\xi_1, \ldots, \xi_t)$ are generated with the given law we generate the branching process as follows: Produce the offspring in the nth generation as a Poisson distribution with the right parameter given by the previous generation (possibly

zero). If there is at least one offspring use ξ_n to decide whether it is a mutant and if so give it fitness F_n. Then use other newly sampled Bernoulli variables with parameter β and fitnesses to decide whether other variables are mutants and if they are decide their fitness. Then surival implies $W_n \geq \xi_n F_n$ as required.

Now $N := \sum_{i=1}^{t} \xi_i$ is binomially distributed with parameters $\beta > 0$ and $t \in \mathbb{N}$. We infer that, for any fixed $x > 1$,

$$\mathbb{P}(W_i \leq x \text{ for all } i \leq t) \leq \mathbb{P}(F_i \xi_i \leq x \text{ for all } i \leq t) + \mathbb{P}(\text{extinction})$$

$$= \sum_{i=0}^{t} \binom{t}{i} \beta^i (1 - \beta)^{t-i} \mathbb{P}(F \leq x)^i + \mathbb{P}(\text{extinction})$$

$$= (\beta \mathbb{P}(F \leq x) + (1 - \beta))^t + \mathbb{P}(\text{extinction}).$$

Since $\mathbb{P}(F \leq x) < 1$ and $\beta > 0$, we get

$$\mathbb{P}(W_i \leq x \text{ for all } i) = \lim_{t \uparrow \infty} \mathbb{P}(W_i \leq x \text{ for all } i \leq t)$$

$$\leq \lim_{t \uparrow \infty} (\beta \mathbb{P}(F \leq x) + (1 - \beta))^t + \mathbb{P}(\text{extinction}) = \mathbb{P}(\text{extinction}),$$

hence $\mathbb{P}((W_t) \text{ is unbounded}) = \mathbb{P}(\text{survival})$ as claimed. $\square$

We next describe the distribution of W_t given the process at time $t - 1$.

Lemma 3 *Suppose that at generation $t - 1$ there are n individuals with fitness F_1, F_2, ..., F_n and set $\mathcal{X} := \sum_{i=1}^{n} F_i$. Then, for all $x \geq 0$,*

$$\mathbb{P}(W_t > x) = 1 - e^{-\beta \mathcal{X} G(x)}.$$

Proof First fix a positive integer n and suppose $W_t^{(n)}$ is the largest of n independently sampled fitnesses and $W_t^{(0)} = 0$. Let $\bar{G}(x) = 1 - G(x)$ and note that

$$\mathbb{P}(W_t^{(n)} > x) = 1 - \mathbb{P}(W_t^{(n)} \leq x) = 1 - \bar{G}(x)^n.$$

Now let N be the number of mutants in generation t, which is Poisson distributed with mean $\beta \mathcal{X}$. Hence, for $x \geq 0$,

$$\mathbb{P}(W_t > x) = \sum_{n=0}^{\infty} \mathbb{P}(W_t^{(n)} > x) \mathbb{P}(N = n) = \sum_{n=0}^{\infty} (1 - \bar{G}(x)^n) \mathbb{P}(N = n)$$

$$= 1 - \sum_{n=0}^{\infty} \bar{G}(x)^n \frac{(\beta \mathcal{X})^n}{n!} e^{-\beta \mathcal{X}} = 1 - e^{-\beta \mathcal{X}(1 - \bar{G}(x))}.$$

As $1 - \bar{G}(x) = G(x)$ the proof is complete. $\square$

The next two results concern the potential limit $v(\alpha)$. We first characterise $v(\alpha)$ as a maximum and then as the growth rate in a recursion relation. Note that the first result easily implies that $v(\alpha)$ is decreasing, as well as continuous and positive.

Lemma 4 *We have*

$$v(\alpha) = \max \left\{ \tfrac{1}{m} \log(m/\alpha) \colon m \in \mathbb{N} \right\}.$$

In particular, for all $m \in \mathbb{N}$, we have

$$m \leq \alpha e^{v(\alpha)m} \quad \text{and} \quad T = \alpha e^{v(\alpha)T}. \tag{8}$$

Proof First recall $(T-1)^T/T^{T-1} < \alpha \leq T^{T+1}/(T+1)^T$ and observe that $\alpha > m^{m+1}/(m+1)^m$ for $m < T$ and $\alpha \leq m^{m+1}/(m+1)^m$ for $m \geq T$, where $m \in \mathbb{N}$. Since

$$m(m+1)\left(\frac{1}{m}\log\frac{m}{\alpha} - \frac{1}{m+1}\log\frac{m+1}{\alpha}\right) = \log\frac{m^{m+1}}{(m+1)^m\alpha},$$

we have the desired result. $\qquad\square$

Remark 5 Let $\alpha_T := T^{T+1}/(T+1)^T$. The equality $m = \alpha e^{\nu(\alpha)m}$ holds iff $(m = T)$ or $(m = T+1$ and $\alpha = \alpha_T)$.

For the remainder of this subsection, we abbreviate $\nu := \nu(\alpha)$.

Lemma 6 *For some positive sequence (a_n) we define inductively*

$$\chi_t := \chi_t(\alpha, (a_n)) := \max\left\{a_t, \frac{t-1}{\alpha}\chi_1, \ldots, \frac{1}{\alpha}\chi_{t-1}\right\}. \tag{9}$$

Then, if $\lim_{n\to\infty} a_n e^{-\nu n} = 0$, there are positive constants c and c' such that

$$c'e^{\nu t} \leq \chi_t \leq ce^{\nu t} \quad \text{for all } t \geq 1,$$

and therefore we have

$$\lim_{t\to\infty}\frac{\log\chi_t}{t} = \nu.$$

Proof Abbreviate $\hat{\chi}_t := \chi_t/c'$ with $c' = e^{-\nu T}\min\{\chi_1, \chi_2, \ldots, \chi_T\}$. Obviously, $\hat{\chi}_t \geq e^{\nu t}$ for $t \leq T$. Now assume $n \geq T$ and $\hat{\chi}_t \geq e^{\nu t}$ for all $t \leq n$. By the assumption and (8), we have

$$\hat{\chi}_{n+1} \geq \frac{T}{\alpha}\hat{\chi}_{n+1-T} \geq \frac{T}{\alpha}e^{\nu(n+1-T)} = e^{\nu(n+1)}.$$

Induction gives $\hat{\chi}_t \geq e^{\nu t}$ and hence $\chi_t \geq c'e^{\nu t}$ for all $t \geq 1$.

Now, choose a positive integer n_0 such that $a_n \leq e^{\nu n}$ for all $n \geq n_0$. Let $\bar{\chi}_t = \chi_t/c$ with $c = \max\{1, \chi_1, \chi_2, \ldots, \chi_{n_0}\}$. Obviously, $\bar{\chi}_t \leq 1 \leq e^{\nu t}$ for all $t \leq n_0$. Now let $n \geq n_0$ and assume that $\bar{\chi}_t \leq e^{\nu t}$ for all $t \leq n$. Then,

$$\bar{\chi}_{n+1} = \max\left\{\frac{a_{n+1}}{c}, \frac{n}{\alpha}\bar{\chi}_1, \frac{n-1}{\alpha}\bar{\chi}_2, \ldots, \frac{n-k+1}{\alpha}\bar{\chi}_k, \ldots, \frac{1}{\alpha}\bar{\chi}_n\right\}$$

$$\leq \max\left\{e^{\nu(n+1)}, \frac{n}{\alpha}e^\nu, \frac{n-1}{\alpha}e^{2\nu}, \ldots \frac{n-k+1}{\alpha}e^{\nu k}, \ldots, \frac{1}{\alpha}e^{\nu n}\right\} \leq e^{\nu(n+1)},$$

where we have used (8). By induction, we have $\chi_t \leq ce^{\nu t}$ for all $t \geq 1$. $\qquad\square$

For later reference we define

$$\tilde{\chi}_i(t) := \begin{cases} -\infty, & \text{if } i < 0, \\ a_t, & \text{if } i = 0, \\ (t-i)\chi_i/\alpha, & \text{if } 1 \leq i \leq t-1. \end{cases} \tag{10}$$

Lemma 7 *Define*

$$I_t := \max\{i < t \colon \chi_t = \tilde{\chi}_i(t)\}. \tag{11}$$

Then $t - I_t$ is bounded.

Proof By Lemma 6 we have $c'e^{vt} \leq \chi_t \leq ce^{vt}$ for all t. Since there is t_0 such that $c'e^{vt} > a_t$ for all $t \geq t_0$, we can write, for $t > t_0$,

$$\chi_t = \max\left\{\tfrac{t-1}{\alpha}\chi_1, \ldots, \tfrac{1}{\alpha}\chi_{t-1}\right\}.$$

Now it is enough to show that $t - I_t$ is bounded for $t > t_0$.

Note that, for $1 \leq m \leq t - 1$,

$$ce^{vt}A(m) \geq \tfrac{m}{\alpha}\chi_{t-m} \geq c'e^{vt}A(m),$$

where $A(m) = me^{-vm}/\alpha$ with $A(T) = 1$. Since $\lim_{m\to\infty} A(m) = 0$, there is m_0 such that $c' > cA(m)$ and hence

$$\tfrac{m}{\alpha}\chi_{t-m} < c'e^{vt}, \quad \text{for all } m \geq m_0.$$

As the right hand side is a lower bound of $\tfrac{T}{\alpha}\chi_{t-T}$ we get that $t - I_t$ cannot be larger than $\max\{m_0, t_0\}$, as desired. $\square$

Remark 8 If we choose $T' > \sup\{t - I_t : t \in \mathbb{N}\}$, then, for all t,

$$\chi_t = \max\{a_t, (t-1)\chi_1/\alpha, \ldots, \chi_{t-1}/\alpha\} = \max\{\tilde{\chi}_{t-T'}(t), \ldots, \tilde{\chi}_{t-1}(t)\}.$$

In words, χ_t is completely determined by $\tilde{\chi}_i(t)$ for i within the window $t - T' \leq i \leq t - 1$. This fact will play an important role in the proof of Theorem 1.

We conclude the subsection with two estimates for classical Galton–Watson processes.

Lemma 9 *Consider a supercritical Galton–Watson process $(\mathcal{X}_t)_{t\geq 0}$ with Poisson offspring distribution with mean $\theta > 1$, starting in generation 0 with a single individual. Fix $0 < x < 1$ and an integer $n \geq 1$. Then,*

$$\mathbb{P}\left(\mathcal{X}_t \geq x^t\theta^t \text{ for all } 1 \leq t \leq n\right) \geq 1 - n\frac{(1-x)^{-2}}{\theta - 1}. \tag{12}$$

Proof First note that the mean and the variance of $\mathcal{X}_t$ are (see, e.g., [37])

$$\mathbb{E}[\mathcal{X}_t] = \theta^t, \quad \mathbb{V}[\mathcal{X}_t] = \frac{\theta^{2t} - \theta^t}{\theta - 1},$$

respectively and that

$$\mathbb{P}\left(\mathcal{X}_t \geq x^t\theta^t \text{ for all } 1 \leq t \leq n\right) \geq 1 - \sum_{t=1}^{n}\mathbb{P}\left(\mathcal{X}_t \leq x^t\theta^t\right), \tag{13}$$

where we have used the sub-additivity of the probability measure. Using Chebyshev's inequality, we get

$$\mathbb{P}\left(\mathcal{X}_t \leq x^t\theta^t\right) = \mathbb{P}\left(\theta^t - \mathcal{X}_t \geq \theta^t - x^t\theta^t\right) \leq \mathbb{P}\left(\left|\mathcal{X}_t - \theta^t\right| \geq \theta^t - x^t\theta^t\right)$$

$$\leq \left(1 - x^t\right)^{-2}\frac{\mathbb{V}[\mathcal{X}_t]}{\theta^{2t}} = \left(1 - x^t\right)^{-2}\frac{1 - \theta^{-t}}{\theta - 1} \leq \frac{(1-x)^{-2}}{\theta - 1},$$

which, along with (13), gives the claimed inequality. $\square$

Lemma 10 *For a Galton–Watson process* $(\mathcal{X}_t)$ *with* $\mathcal{X}_0 = K_0$ *and generation dependent offspring distribution* N_t *with* $\mathbb{E}[N_t] \leq N$ *for all t,*

$$\mathbb{P}\big(\mathcal{X}_t \leq KN^t B^t \text{ for all } t \geq 1\big) \geq 1 - \frac{K_0}{K(B-1)},$$

for all $B > 1$ *and* $K > 0$.

Proof By Markov's inequality, we have

$$\mathbb{P}\left(\mathcal{X}_t \geq KN^t B^t\right) \leq \frac{\mathbb{E}[\mathcal{X}_t]}{KN^t} B^{-t}.$$

Since $\mathbb{E}[\mathcal{X}_{t+1}|\mathcal{X}_t] = \mathcal{X}_t \mathbb{E}[N_t]$, we have $\mathbb{E}[\mathcal{X}_t] = K_0 \prod_{i=0}^{t-1} \mathbb{E}[N_i] \leq K_0 N^t$, which gives

$$\mathbb{P}\left(\mathcal{X}_t \geq KN^t B^t\right) \leq \frac{K_0}{KB^t}.$$

Since

$$\mathbb{P}(\mathcal{X}_t \leq KN^t B^t \text{ for all } t \geq 1) \geq 1 - \sum_{t=1}^{\infty} \mathbb{P}(\mathcal{X}_t \geq KN^t B^t),$$

a geometric sum gives the claimed inequality. □

4.2 Proof of the Lower Bound

In this subsection we show that, for given $\alpha > 0$ and all $\alpha' > \alpha$, we have

$$\mathbb{P}\left(\liminf_{t\to\infty} \frac{\log\log X(t)}{t} \geq \nu(\alpha') \,\Big|\, \text{survival}\right) = 1. \tag{14}$$

In both models at each generation s, we can regard a lineage originating from the mutant with fitness W_s as a version $(\hat{X}_t(f))_{t\geq s}$ of the same model starting in generation s with a single individual of fitness $f = W_s$. Since $X(t) \geq \hat{X}_t(f)$ for $t \geq s$, (14) is proved, if there is at least one s such that

$$\liminf_{t\to\infty} \frac{\log\log \hat{X}_t(W_s)}{t} \geq \nu(\alpha').$$

As (W_t) is unbounded almost surely on survival it therefore suffices to show that

$$\lim_{f\to\infty} \mathbb{P}\left(\liminf_{t\to\infty} \frac{\log\log \hat{X}_t(f)}{t} \geq \nu(\alpha')\right) = 1. \tag{15}$$

For convenience, we use the convention $(\log\log \hat{X}_t(f))/t = -\infty$ if $\hat{X}_t(f) = 0$. As $(\hat{X}_t(x))$ can be coupled to an FMM $(S_t(x))$ with the same initial condition such that $\hat{X}_t(x) \geq S_t(x)$ for all $t \geq 0$, the result follows by combining Lemma 6 with the following statement.

Lemma 11 *Fix* $0 < \epsilon < 1/2$ *and let*

$$E(f) := \big\{S_t(f) \geq (1-\beta)^t f^{\chi_t'} \text{ for all } t \geq 1\big\},$$

where $\chi_t' := \chi_t(\alpha', (\frac{n}{2}))$ *with* $\alpha' := \alpha/(1-2\epsilon)$. *Then*

$$\lim_{f\to\infty} \mathbb{P}(E(f)) = 1.$$

Proof We define $m_0 := f$, $n_0 := f^{1/2}$, and $(i \geq 1)$

$$m_i := f^{(1-\epsilon)\chi_i'/\alpha}, \quad n_i := f^{(1-2\epsilon)\chi_i'/\alpha} = f^{\chi_i'/\alpha'}.$$

For later reference, we also define $U_i := n_i/m_i$ for all $i \geq 0$.

Set $\epsilon_0 = \epsilon/(2 - 2\epsilon)$. By our assumption on μ, there is f_0 such that

$$G(x) \geq x^{-\alpha(1+\epsilon_0)} \text{ for all } x > f_0.$$

Since we are only interested in the limit as $f \to \infty$, we may assume that f is so large that $(1 - \beta)m_0 \geq 2$, $(1 - \beta)m_1 \geq 2$, $U_0 < 1/2$, $U_1 < 1/2$, and $m_1 > f_0$. Notice that by assumption,

$$G(m_1) \geq G(m_i) \geq m_i^{-\alpha(1+\epsilon_0)} = f^{-(1-\epsilon/2)\chi_i'} \text{ for all } i \geq 1.$$

For χ_t', we choose T' as in Remark 8. By $N_{i,t}$ we denote the number of individuals with fitness W_i at generation t. Define events

$$A_i := \{N_{i,t} \geq (1 - \beta)^{t-i} n_i^{t-i} \text{ for all } i < t \leq i + T'\}, \quad B_i := \{W_i \geq m_i\}.$$

Let D_{-1} be the certain event and, for $i \in \mathbb{N}_0$,

$$D_i := D_{i-1} \cap A_i \cap B_i, \quad A := \bigcap_{i=0}^{\infty} D_i.$$

Now observe that

$$\mathbb{P}(A) = \lim_{i \to \infty} \mathbb{P}(D_i), \quad \mathbb{P}(D_i) = \mathbb{P}(A_i|B_i \cap D_{i-1})\mathbb{P}(B_i|D_{i-1})\mathbb{P}(D_{i-1}).$$

By Lemma 9 we have

$$\mathbb{P}(A_i|B_i \cap D_{i-1}) \geq 1 - \frac{T'(1 - n_i/W_i)^{-2}}{(1 - \beta)W_i - 1} \geq 1 - \frac{T'(1 - U_i)^{-2}}{(1 - \beta)m_i - 1},$$

where we have used (12).

To proceed, we find the $\mathcal{X}$ in Lemma 3 on the event D_{n-1} as

$$\mathcal{X} \geq \sum_{i=0}^{t-1} N_{i,t-1} W_i \geq (1 - \beta)^t \sum_{i=t-T'-1}^{t-1} f^{\tilde{\chi}_i'(t)} \geq (1 - \beta)^t f^{\chi_t'},$$

where we have used $W_i \geq m_i \geq n_i$ and $\tilde{\chi}_i'(t)$ as in (10) for parameters α' and $a_s = s/2$. Using Lemma 3 with $G(m_i) \geq f^{-(1-\epsilon/2)\chi_i'}$, we have

$$\mathbb{P}(B_i|D_{i-1}) \geq 1 - \exp\left(-\beta(1 - \beta)^i f^{\epsilon\chi_i'/2}\right).$$

Now we define

$$b_i := \frac{T'(1 - U_i)^{-2}}{(1 - \beta)m_i - 1} + (1 - \delta_{i0}) \exp\left(-\beta(1 - \beta)^i f^{\epsilon\chi_i'/2}\right), \quad \phi(f) := \sum_{i=0}^{\infty} b_i,$$

where δ_{ij} is the Kronecker delta symbol. Trivially, we have $\lim_{f \to \infty} b_i = 0$ for all $i \geq 0$. Since, for sufficiently large f, b_s for fixed s is a bounded and decreasing function of f and since Lemma 6 gives

$$\lim_{s \to \infty} b_s 2^s = 0,$$

there is s_0 such that $|b_s| < 2^{-s}$ for all $s > s_0$ and for all assumed value of f. Therefore, the series defining $\phi(f)$ converges uniformly for sufficiently large f and $\lim_{f \to \infty} \phi(f) = 0$. Therefore, for sufficiently large f, we get

$$\mathbb{P}(A) \geq \prod_{i=0}^{\infty}(1 - b_i) \geq 1 - \phi(f), \quad \lim_{f \to \infty} \mathbb{P}(A) = 1,$$

where we have used $(1 - x)(1 - y) \geq 1 - x - y$ for $x, y \geq 0$. As, on the event A,

$$S_t(f) \geq \sum_{i=t-T'}^{t} N_{i,t} \geq (1 - \beta)^t f^{\chi_t'}$$

where we have assumed $N_{i,t} = 0$ for $i < 0$, we see that $A \subset E(f)$ and the proof is completed. □

In fact, Lemma 11 and its proof are applicable to the MMM verbatim, except that S_t is replaced by $\hat{X}_t$. If we are interested in the proof only for the MMM, we actually do not need to introduce S_t.

4.3 Proof of the Upper Bound

In this subsection we show that, for given $\alpha > 0$ and all $0 < \alpha' < \alpha$, we have for the MMM denoted by (M_t), or $(M_t(x))$ if in the initial generation there is a single individual with fixed fitness x, that

$$\mathbb{P}\left(\limsup_{t \to \infty} \frac{\log \log M_t}{t} \leq \nu(\alpha') \right) = 1. \tag{16}$$

In case of extinction the upper bound holds by convention. One can construct two processes with initial fitness $x \leq y$ on the same probability space such that $M_t(x) \leq M_t(y)$ for all t. Indeed, this can be done as follows. First construct $(M_t(y))$ and look at its genealogical tree truncated after the first mutant in every line of descent from the root. Removing any individual in that tree together with all its offspring from $(M_t(y))$ independently with probability x/y we obtain $(M_t(x))$.

We now construct an MMM with special initial conditions. Fix $\epsilon > 0$. For given α, let $\delta = (1 + 2\epsilon)/(1 + 3\epsilon)$, $\alpha' = \alpha/(1 + 3\epsilon)$, $\nu' = \nu(\alpha')$, $T = T(\alpha')$, and

$$\Delta_0 = \frac{\epsilon}{\nu'(1 + 3\epsilon)},$$

which is equivalent to $\nu'\Delta_0 + \delta = 1$. We choose Δ such that $0 < \Delta \leq \Delta_0$ and

$$\hat{n} := \frac{T}{\Delta}$$

is an integer. We define, for a given $f > 0$,

$$\chi_t' := e^{\nu' t}, \quad \kappa_{n,t} := e^{\nu'(t-T+n\Delta)}, \quad g_{n,t} := f^{\kappa_{n,t}/\alpha}, \quad h_{n,t} := f^{(1+\epsilon)\kappa_{n,t}/\alpha}.$$

Note that χ_t' above is different from that in Lemma 11.

We briefly explain the motivation of introducing $h_{n,t}$ and other quantities to find an upper bound. Unlike the proof for the lower bound in Lemma 11, where we have only to investigate a single lineage $\hat{X}_t$, we have to consider all mutants to find an appropriate upper bound of the MMM. Since we only need an inequality for the proof, we divide the fitness space of mutants

by $h_{n,t}$ and regard mutants appearing at generation t with fitness in the region $(h_{n,t}, h_{n+1,t}]$ as a mutant class with growth rate bounded by $h_{n+1,t}$.

We consider the MMM $(\tilde{M}_t(f))_{t \geq T-1}$ starting in generation $T-1$ with an initial condition such that there are T different mutant classes with fitness $g_{\hat{n},m}$ for $0 \leq m \leq T-1$ and the number of individuals with fitness $g_{\hat{n},m}$ is $\lfloor (g_{\hat{n},m})^{T-m-1} \rfloor$. We only consider f sufficiently large so that $(1-\beta)f > 2$ and $(1-\beta)f^{\epsilon/\alpha} > 2$.

Now assume that we have proved, for all $\alpha' < \alpha$,

$$\lim_{f \to \infty} \mathbb{P} \left(\limsup_{t \to \infty} \frac{\log \log \tilde{M}_t(f)}{t} \leq \nu(\alpha') \right) = 1. \tag{17}$$

Given an arbitrary $f > 0$ and $\varepsilon > 0$ pick f_ε such that the probability above exceeds $1 - \varepsilon$ and the smallest fitness in the initial condition of $\tilde{M}_t(f_\varepsilon)$ is larger than f. Then (17) guarantees that

$$\mathbb{P} \left(\limsup_{t \to \infty} \frac{\log \log M_t(f)}{t} \leq \nu(\alpha') \right) > 1 - \varepsilon,$$

which proves (16). So it is enough to prove (17). Once (17) is proved, we use the natural coupling such that $S_t \leq M_t$ for all t. Then, almost surely on survival,

$$\limsup_{t \to \infty} \frac{\log \log S_t}{t} \leq \limsup_{t \to \infty} \frac{\log \log M_t}{t} \leq \nu(\alpha),$$

which completes the proof of Theorem 1.

Lemma 12 *In an MMM, let Z_t be the number of* non-mutated *descendants at generation t of $\mathcal{X}$ individuals at generation m whose fitness values are within a bounded interval I with right endpoint b. Assume $\mathcal{X} \leq K$. Then, for all $B > 1$,*

$$\mathbb{P}(Z_t \leq K(1-\beta)^{t-m} b^{t-m} B^{t-m} \text{ for all } t \geq m+1) \geq 1 - \frac{1}{B-1}.$$

Proof As the mean number of non-mutated offspring of an individual is bounded by $(1-\beta)b$ we get the result by applying Lemma 10. □

Lemma 13 *Suppose at generation $t-1$ of an MMM the population consists of n individuals with fitness $F_1, \ldots, F_n$. Let*

$$Y_t := \beta \sum_{i=1}^{n} F_i \tag{18}$$

and let Z be the number of mutants in generation t with fitness in the interval $(a, b]$. Then, with $p := \mu((a, b])$, we have

$$\mathbb{P}(Z > K) = e^{-Y_t p} \sum_{n=K+1}^{\infty} \frac{(Y_t p)^n}{n!}.$$

Proof Observe that Z is Poisson distributed with mean $Y_t p$. □

Remark 14 Using Markov's inequality, we get

$$\mathbb{P}(Z > K) \leq \frac{Y_t p}{K}, \tag{19}$$

which is useful when $K \gg Y_t p$. By Chebyshev's inequality, for $K > Y_t$,

$$\mathbb{P}(Z > K) \leq \mathbb{P}(|Z - Y_t p| \geq K - Y_t p) \leq \frac{Y_t p}{(K - Y_t p)^2} \leq \frac{Y_t}{(K - Y_t)^2}, \tag{20}$$

which is useful when $(K - Y_t)^2 \gg Y_t$. For $K = 0$, we will use

$$\mathbb{P}(Z = 0) = e^{-Y_t p} \geq 1 - Y_t p \geq 1 - Y_t G(a). \tag{21}$$

We denote the number of non-mutated descendants at generation $t \geq T - 1$ of initial individuals with fitness $g_{\hat{n},m}$ by $N_{m,T-1,t}$ and define

$$N_{T-1,t} := \sum_{m=0}^{T-1} N_{m,T-1,t}.$$

The number of mutants that appear at generation $t \geq T$ with fitness in the interval $(h_{n-1,t}, h_{n,t}]$ is denoted by $N_{n,t,t}$ for $0 \leq n \leq \hat{n} + 1$, where we have assumed $h_{-1,t} := 0$ and $h_{\hat{n}+1,t} := \infty$. Typically, $N_{\hat{n}+1,t,t}$ will be zero. The number of non-mutated descendants of $N_{n,m,m}$ at generation $t > m$ is denoted by $N_{n,m,t}$. For $t \geq m \geq T$ define

$$N_{m,t} := \sum_{n=0}^{\hat{n}+1} N_{n,m,t},$$

which gives

$$\tilde{M}_t(f) = \sum_{m=T-1}^{t} N_{m,t}.$$

Let $(\theta_t)_{t \geq T-1}$ be a sequence satisfying $\theta_{T-1} = T$ and, for $t \geq T$,

$$\theta_t = (t - T)\hat{n} + \sum_{m=T-1}^{t-1} \theta_m.$$

Since $\theta_{t+1} - \theta_t = \theta_t + \hat{n}$ for $t \geq T$, we have

$$\theta_t = \begin{cases} 2^{t-T}(T + \hat{n}) - \hat{n}, & \text{for } t \geq T \\ T, & \text{for } t = T - 1. \end{cases} \tag{22}$$

Lemma 15 *For $T \leq x \leq m < t$ (t, m are integers and x is real), we have*

$$e^{\nu'm} + \frac{t-m}{\alpha'} e^{\nu'(m-T)} \leq e^{\nu't}, \tag{23}$$

$$e^{\nu'm} - e^{\nu'(x-\Delta)} + \delta \frac{t-m}{\alpha'} e^{\nu'x} \leq e^{\nu't}. \tag{24}$$

Proof Using (8), we have

$$e^{\nu'm}\left(1 + \frac{t-m}{\alpha'} e^{-\nu'T}\right) = e^{\nu'm} \frac{T+t-m}{T} \leq e^{\nu't} e^{\nu'T} \frac{\alpha'}{T} = e^{\nu't},$$

which proves (23). If $\delta(t - m) - \alpha' e^{-\nu'\Delta}$ is negative, then (24) is trivially valid. If $\delta(t - m) - \alpha' e^{-\nu'\Delta}$ is positive, then the left hand side of (24) has maximum at $x = m$. Therefore, it is enough to prove (24) only for $x = m$. Plugging $x = m$, we have

$$e^{v'm} - e^{v'(m-\Delta)} + \delta \frac{t-m}{\alpha'} e^{v'm} = e^{v'm}(1 - e^{-v'\Delta}) + \delta \frac{t-m}{\alpha'} e^{v'm}$$
$$\leq e^{v't}(v'\Delta + \delta) \leq e^{v't},$$

where we have used $1 - e^{-x} \leq x$, $e^{v'm} \leq e^{v't}$, and $t - m \leq \alpha' e^{v'(t-m)}$. $\qquad\square$

Lemma 16 *Let* $E(f) := \{\tilde{M}_t(f) \leq \theta_{t+1} f^{\chi_t'} \text{ for all } t \geq T\}$. *Then*

$$\lim_{f \to \infty} \mathbb{P}(E(f)) = 1,$$

which implies (17).

Proof Set $\epsilon_0 = \epsilon/(2 + 2\epsilon)$. By our assumption on μ there is f_0 such that $G(f) \leq f^{-\alpha(1-\epsilon_0)}$ for all $f > f_0$. Now we assume $h_{0,T} = f^{(1+\epsilon)/\alpha} > f_0$, which gives

$$G(h_{n,t}) \leq f^{-(1+\epsilon/2)\kappa_{n,t}}, \quad \text{for all } 0 \leq n \leq \hat{n} \text{ and } t \geq T. \tag{25}$$

Let

$$A_{n,T-1} := \left\{ N_{n,T-1,t} \leq \left\lfloor (g_{\hat{n},n})^{T-n-1} \right\rfloor (1-\beta)^{t-T+1} f^{(t-T+1)\chi_n'/\alpha'} \text{ for all } t \geq T \right\}$$

and define $A_{T-1} := \bigcap_{n=0}^{T-1} A_{n,T-1}$. By Lemma 12 with $K = \lfloor (g_{\hat{n},n})^{T-n-1} \rfloor$, $b = g_{\hat{n},n} = f^{\chi_n'/\alpha}$, and $bB = f^{\chi_n'/\alpha'} = f^{(1+3\epsilon)\chi_n'/\alpha}$, we have

$$1 - \mathbb{P}(A_{n,T-1}) \leq \left(f^{3\epsilon\chi_n'/\alpha} - 1 \right)^{-1}, \quad \mathbb{P}(A_{T-1}) \geq 1 - b_{T-1},$$

$$b_{T-1} := \sum_{n=0}^{T-1} \left(f^{3\epsilon\chi_n'/\alpha} - 1 \right)^{-1}.$$

Since $me^{-v'm} \leq \alpha'$ and $g_{\hat{n},n} \leq f^{\chi_n'/\alpha'}$, we have

$$N_{n,T-1,t} \leq f^{\chi_n'(t-n)/\alpha'} \leq f^{\chi_t'} \quad \text{for all } n, t.$$

Therefore,

$$\sum_{n=0}^{T-1} N_{n,T-1,t} \leq T f^{\chi_t'}, \quad \text{for all } t \geq T, \tag{26}$$

on the event A_{T-1}. Let

$$A_{n,m,t} := \begin{cases} \left\{ N_{0,m,t} \leq \theta_m f^{\chi_m'}(1-\beta)^{t-m} f^{(t-m)\kappa_{0,m}/\alpha'} \right\}, & \text{for } n = 0, \\ \left\{ N_{n,m,t} \leq \frac{f^{\chi_m'}}{(g_{n-1,m})^\alpha}(1-\beta)^{t-m} f^{\delta(t-m)\kappa_{n,m}/\alpha'} \right\}, & \text{for } 1 \leq n \leq \hat{n}, \\ \left\{ N_{\hat{n}+1,m,t} = 0 \right\} & \text{for } n = \hat{n} + 1. \end{cases}$$

Note that $A_{n,t,t}$ has information on the empirical distribution of mutants' fitness that appear at generation t. We define

$$\tilde{A}_{n,m} = \bigcap_{t=m+1}^{\infty} A_{n,m,t}, \quad A_{n,m} := A_{n,m,m} \cap \tilde{A}_{n,m}, \quad A_m := \bigcap_{n=0}^{\hat{n}+1} A_{n,m},$$

$$D_{T-1} := A_{T-1}, \quad D_n := D_{n-1} \cap A_n, \quad A(f) := \bigcap_{n=T-1}^{\infty} D_n.$$

By Lemma 15, we have on the event A_m, that for all $t \geq m \geq T$,

$$N_{m,t} = \sum_{n=0}^{\hat{n}+1} N_{n,m,t} \leq (\theta_m + \hat{n}) f^{\chi'_t},$$

and, in turn,

$$\tilde{M}_t(f) \leq \left((t+1-T)\hat{n} + \sum_{m=T-1}^{t} \theta_m \right) f^{\chi'_t} = \theta_{t+1} f^{\chi'_t}$$

on the event $A(f)$. Therefore, $A(f) \subset E(f)$ and the proof is complete if we show

$$\lim_{f \to \infty} \mathbb{P}(A(f)) = 1.$$

Now we investigate $\mathbb{P}(A_m | D_{m-1})$. First note that

$$1 - \mathbb{P}(A_m | D_{m-1}) \leq \sum_{n=0}^{\hat{n}+1} \left[1 - \mathbb{P}(A_{n,m} | D_{m-1}) \right],$$

$$\mathbb{P}(A_{n,m} | D_{m-1}) = \mathbb{P}(\tilde{A}_{n,m} | A_{n,m,m} \cap D_{m-1}) \mathbb{P}(A_{n,m,m} | D_{m-1}),$$

and, on the event D_{m-1},

$$Y_m \leq \beta \theta_m f^{\chi'_m}, \tag{27}$$

where Y_m is defined in (18).

We begin with $\mathbb{P}(A_{\hat{n}+1,m} | D_{m-1})$, which clearly equals $\mathbb{P}(A_{\hat{n}+1,m,m} | D_{m-1})$. Using (21) with $a = h_{\hat{n},m}$ and $Y_m G(a) \leq \theta_m f^{-\epsilon \chi'_m/2}$, we obtain

$$\mathbb{P}(A_{\hat{n}+1,m} | D_{m-1}) \geq 1 - b_{\hat{n}+1,m}, \quad b_{\hat{n}+1,m} := \theta_m f^{-\epsilon \chi'_m/2}.$$

Now we consider $\mathbb{P}(A_{0,m} | D_{m-1})$. Using (20) with $K = \theta_m f^{\chi'_m}$ and (27), we have

$$\mathbb{P}(A_{0,m,m} | D_{m-1}) \geq 1 - \frac{\beta}{(1-\beta)^2 \theta_m} f^{-\chi'_m}.$$

Using Lemma 12 with $B = f^{\kappa_{0,m}/\alpha'}/h_{0,m} = f^{2\epsilon\kappa_{0,m}/\alpha}$, we have

$$\mathbb{P}(\tilde{A}_{0,m} | A_{0,m,m} \cap D_{m-1}) \geq 1 - \left(f^{2\epsilon\kappa_{0,m}/\alpha} - 1 \right)^{-1}.$$

Therefore, defining

$$b_{0,m} := \frac{\beta}{(1-\beta)^2 \theta_m} f^{-\chi'_m} + \left(f^{2\epsilon\kappa_{0,m}/\alpha} - 1 \right)^{-1},$$

we have $\mathbb{P}(A_{0,m} | D_{m-1}) \geq 1 - b_{0,m}$.

Finally, we move on to $\mathbb{P}(A_{n,m} | D_{m-1})$ for $1 \leq n \leq \hat{n}$. Using (19) with $K = f^{\chi'_m - \kappa_{n-1,m}}$, $a = h_{n-1,m}$, and $Y_m G(a) \leq \theta_m f^{\chi'_m - (1+\epsilon/2)\kappa_{n-1,m}}$, we have

$$\mathbb{P}(A_{n,m,m} | D_{m-1}) \geq 1 - \theta_m f^{-\epsilon\kappa_{n-1,m}/2}.$$

Using Lemma 12 with $B = f^{\delta\kappa_{n,m}/\alpha'}/h_{n,m} = f^{\epsilon\kappa_{n,m}/\alpha}$, we have

$$\mathbb{P}\left(\tilde{A}_{n,m} | A_{n,m,m} \cap D_{m-1} \right) \geq 1 - \left(f^{\epsilon\kappa_{n,m}/\alpha} - 1 \right)^{-1}.$$

Therefore, defining

$$b_{n,m} := \theta_m f^{-\epsilon \kappa_{n-1,m}/2} + \left(f^{\epsilon \kappa_{n,m}/\alpha} - 1 \right)^{-1},$$

we have $\mathbb{P}(A_{n,m}|D_{m-1}) \geq 1 - b_{n,m}$. We define

$$b_m := \sum_{n=0}^{\hat{n}+1} b_{n,m} \text{ for } m \geq T, \quad \text{and } \phi(f) := \sum_{m=T-1}^{\infty} b_m.$$

Recall that we have assumed $(1-\beta)f > 2$ and $(1-\beta)f^{\epsilon/\alpha} > 2$. Since b_m for given m is a bounded function of f which is decreasing to zero and

$$\lim_{m \to \infty} b_m 2^m = 0,$$

there is m_0 such that $|b_m| < 2^{-m}$ for all $m > m_0$. Hence the series defining $\phi(f)$ converges uniformly for sufficiently large f and, accordingly, $\lim_{f \to \infty} \phi(f) = 0$. Therefore, for sufficiently large f,

$$\mathbb{P}(A(f)) \geq \prod_{m=T-1}^{\infty} (1 - b_m) \geq 1 - \phi(f),$$

and $\lim_{f \to \infty} \mathbb{P}(A(f)) = 1$, which completes the proof. $\square$

5 Empirical Fitness Distributions of the FMM

Apart from the fact that the population is dominated by a single mutant class at all times, the proof of the double-exponential growth rate ν presented in Sect. 4 does not give any insight into the structure of the population. However, since the solution χ_t of the recursion relation (9) correctly describes the asymptotic growth of $X(t)$, it provides a natural starting point for addressing this question at least on a heuristic level. In this section, we analyze the recursion relation in more depth to understand the demographic structure of the FMM in the long time limit, which turns out to display a rather rich behaviour.

5.1 Numerical Solution of the Recursion Relation

To characterize the empirical fitness distribution we introduce the following quantities:

$$J_i(t) := \frac{\log W_i}{\log W_t} \approx \frac{\chi_i}{\chi_t}, \quad P(t) := \frac{\log X(t+1) - \log X(t)}{\log W_t} \approx \alpha \frac{\chi_{t+1} - \chi_t}{\chi_t},$$

$$R_i(t) := \frac{\log W_i^{t-i} - \log X(t)}{\log X(t)} = \frac{t-i}{\alpha} J_i(t) - 1,$$

where the second approximate relations in the definitions of $J_i(t)$ and $P(t)$ become equalities in the formal deterministic limit $f \to \infty$ (see Sect. 3). The ratio $J_i(t) \in [0, 1]$ compares the log-fitness of the mutant class born at time i to the log-fitness of the current fittest mutant. Since $X(t+1) \approx (1-\beta)X(t)\bar{F}_t$ with $\bar{F}_t$ denoting the mean fitness of the population at generation t, $P(t)$ quantifies the mean fitness at generation t on the same scale. The decomposition in Eq. (3) shows that the fraction of the population in mutant class i at time t
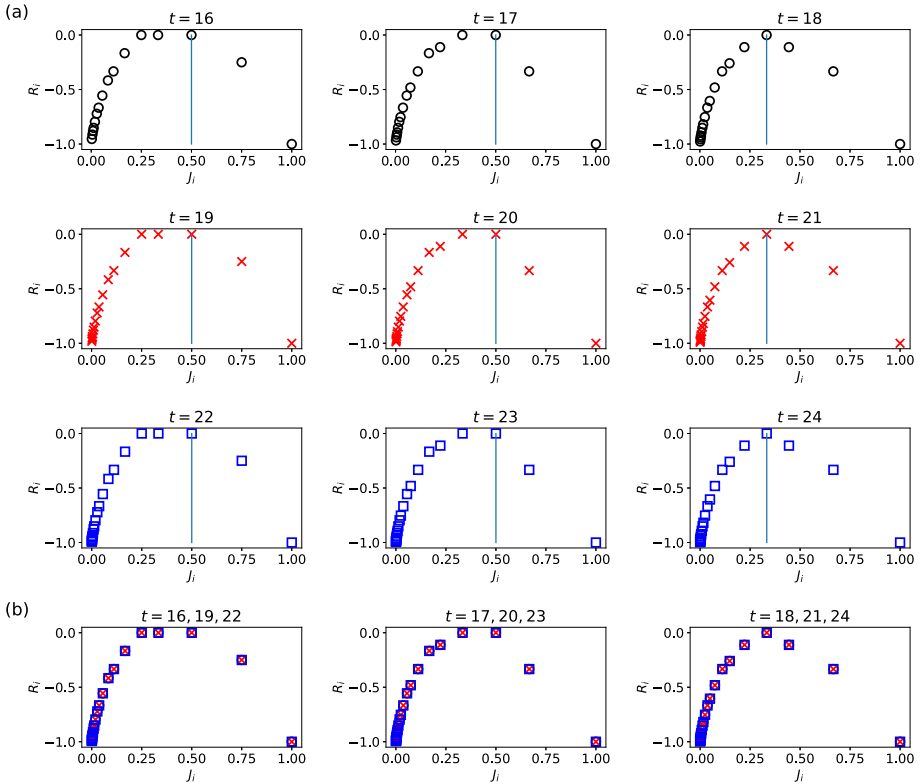
(a)

(b)

Fig. 2 **a** Plots of $R_i(t)$ vs. $J_i(t)$ for $\alpha = 1$ at generations $t = 16, 17, \ldots, 24$. The vertical line at each panel indicates the location of the mean fitness $P(t)$. **b** Data collapse plots of $R_i(t)$ vs. $J_i(t)$ for each column of **a**

is proportional to W_i^{t-i}, and therefore $R_i(t) \in [-1, 0]$ serves as proxy of the (logarithmic) empirical fitness distribution at generation t over mutant classes i.

In Fig. 2 a, we plot $R_i(t)$ against $J_i(t)$ for nine consecutive generations, obtained by numerically solving the recursion relation (5) for $\alpha = 1$ with $a_n = n$. The salient feature is the periodic behavior of the fitness distribution with period 3; note that $T = 3$ for $\alpha = 1$. To illustrate the accuracy of the periodic behaviour, we present data-collapse plots in Fig. 2b. In most regions of J_i, the collapse looks perfect (since the number of mutant classes increases with the number of generations, the empirical fitness distributions at different times cannot be identical). For another illustration of the periodicity, we depict $P(t)$ vs. t for various values of α (with $a_n = n$) in Fig. 3. After an early transient behavior, $P(t)$ clearly shows periodic behavior. A rigorous proof of the periodicity will be given in Sect. 5.2.

The periodicity was taken into account in the numerical estimates of ν reported in Fig. 1. Rather than monitoring $\log \chi_t / t$, which converges very slowly, we computed the quantity

$$\hat{\nu}(t) := \frac{1}{T} \log \frac{\chi_{t+T}}{\chi_t},$$

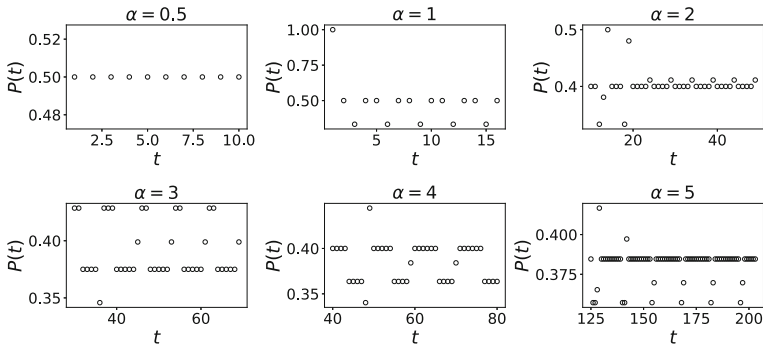which approaches a constant in a relatively short time.

Fig. 3 Plots of $P(t)$ vs. t for various α with $a_n = n$. The function $P(t)$ exhibits complete periodicity for large t

5.2 Periodicity of $\chi_t e^{-\nu t}$

By Lemma 6, we know that

$$c_t := \chi_t e^{-\nu t}$$

is bounded away from zero and infinity. Now we show that c_t is not only bounded, but eventually becomes periodic.

Proposition 17 *For any sequence (a_n) in the recursion relation (9), there is a t_1 such that $c_t = c_{t+T}$ for all $t \geq t_1$.*

Proof In this proof, k and k' are exclusively used as integers in the range $1 \leq k, k' \leq T$. Since $\chi_{t+T} \geq e^{\nu T} \chi_t$ (see Sect. 3), the sequence $(c_{k+nT})_n$ is nondecreasing and bounded. Consequently,

$$C_k := \lim_{n \to \infty} c_{k+nT} \tag{28}$$

is well defined. Note that $\max\{C_k : 1 \leq k \leq T\}$ becomes the optimal upper bound in Lemma 6. If n satisfies $nT > T'$ with $T' > \max\{t - I_t\}$ (see Remark 8), then we have

$$
\begin{aligned}
c_{k+nT} &= e^{-\nu(k+nT)} \max\left\{ \frac{k+nT-t'}{\alpha} \chi_{t'} : k+nT-T' \leq t' < k+nT \right\} \\
&= \max\left\{ \frac{s}{\alpha} e^{-\nu s} c_{k+nT-s} : 1 \leq s \leq T' \right\}.
\end{aligned}
\tag{29}
$$

Taking n to infinity, we get

$$C_k = \max\left\{ \frac{s}{\alpha} e^{-\nu s} C_{k-s} : 1 \leq s \leq T' \right\}, \tag{30}$$

and, by definition, $C_{k+mT} = C_k$ for any integer m. Since $T e^{-\nu T}/\alpha = 1$ and $C_{k-T} = C_k$, we can rewrite (30) as

$$C_k = \max\left\{ \frac{s}{\alpha} e^{-\nu s} C_{k-s} : 1 \leq s \leq T+1 \right\}.$$

Comparing terms with $s = T-1, T, T+1$ for any k, we have

$$\frac{T-1}{T} e^{\nu} C_{k+1} = \frac{T-1}{\alpha} e^{-\nu(T-1)} C_{k+1} \leq C_k,$$

$$\frac{T+1}{T}e^{-\nu}C_{k-1} = \frac{T+1}{\alpha}e^{-\nu(T+1)}C_{k-1} \le C_k,$$

which gives

$$\frac{T+1}{T} \le \frac{C_{k+1}e^{\nu}}{C_k} \le \frac{T}{T-1},$$

for all k. Let $\varphi_k = C_k e^{\nu}/C_{k-1}$ with $\varphi_1 = \varphi_{T+1} = C_1 e^{\nu}/C_T$. Then,

$$C_k = C_1 e^{-\nu(k-1)} \prod_{j=2}^{k} \varphi_j,$$

for $k > 1$. Setting $k = T+1$ and considering $C_{T+1} = C_1$, we have

$$\prod_{j=1}^{T} \varphi_j = e^{\nu T} = \frac{T}{\alpha}.$$

To sum up, C_k takes the form

$$C_k = C_0 e^{-\nu k} \prod_{j=1}^{k} \varphi_j,$$

where C_0 is a positive constant (note that $\chi_t(\alpha, (C_0 a_n)) = C_0 \chi_t(\alpha, (a_n))$) and φ_k satisfies

$$\frac{T+1}{T} \le \varphi_k \le \frac{T}{T-1}, \quad \prod_{j=1}^{T} \varphi_j = \frac{T}{\alpha} = e^{\nu T}. \tag{31}$$

If $\alpha = \alpha_T$ (see Remark 5), then $e^{\nu} = (T+1)/T$ and the only possible value of φ_k is $\varphi_k = e^{\nu}$ for all k because of (31).

To simplify (29) for large n, we use the following observation. For $p \in \mathbb{N}$ with $X := 1/T$ and $C_{k-(T\pm p)} = C_{k\mp p}$, we have

$$\frac{T-p}{\alpha}e^{-\nu(T-p)}C_{k+p} = \frac{T-p}{T}e^{\nu p}\frac{C_{k+p}}{C_k}C_k = C_k \frac{T-p}{T}\prod_{j=1}^{p}e^{\nu}\frac{C_{k+j}}{C_{k+j-1}}$$

$$\le C_k \frac{T-p}{T}\left(\frac{T}{T-1}\right)^p = C_k \frac{1-pX}{(1-X)^p},$$

and

$$\frac{T+p}{\alpha}e^{-\nu(T+p)}C_{k-p} = \frac{T+p}{T}e^{-\nu p}\frac{C_{k-p}}{C_k}C_k = C_k \frac{T+p}{T}\prod_{j=1}^{p}\frac{C_{k-j}}{e^{\nu}C_{k-j+1}}$$

$$\le C_k \frac{T+p}{T}\left(\frac{T}{T+1}\right)^p = C_k \frac{1+pX}{(1+X)^p}.$$

Since $\sup_{p\ge 2}(1+pX)/(1+X)^p < 1$ for all nonzero X not smaller than -1, relating s and p by $p = |s-T|$ we can choose $\epsilon > 0$ such that $C_k - \epsilon > se^{-\nu s}C_{k-s}/\alpha$ for all s with $|s-T| > 1$ and for all k. By (28), for this ϵ, there is an integer m_0 such that $C_k - \epsilon < c_{k+nT} \le C_k$ for all $n \ge m_0$ and for all k. If $n > m_0$, then

$$c_{k+nT} > C_k - \epsilon > se^{-\nu s}C_{k-s}/\alpha \ge se^{-\nu s}c_{k-s+nT}/\alpha,$$

for all s with $|s - T| > 1$, which reduces (29) to

$$c_{k+(n+1)T} = \max\left\{c_{k+nT}, \tfrac{T+1}{T}e^{-\nu}c_{k-1+nT}, \tfrac{T-1}{T}e^{\nu}c_{k+1+nT}\right\}. \tag{32}$$

In the following, n is assumed so large that (32) is valid for all k. Defining $\delta_{k,n} := 1 - c_{k+nT}/C_k$ with the convention $\delta_{k+mT,n} := \delta_{k,n+m}$ for integer m and using the definition of φ_k, we can write

$$\delta_{k,n+1} = \min\left\{\delta_{k,n}, 1 - \tfrac{T+1}{T\varphi_k}(1 - \delta_{k-1,n}), 1 - \tfrac{T-1}{T}\varphi_{k+1}(1 - \delta_{k+1,n})\right\}.$$

As $c_{k+nT} \to C_k$, we have $\delta_{k,n} \to 0$ as $n \to \infty$. If $(T + 1)/(T\varphi_k) < 1$, then $1 - (T + 1)/(T\varphi_k)(1 - \delta_{k-1,n}) > 1 - (T + 1)/(T\varphi_k) > 0$ for sufficiently large n and, therefore, the term with $(T + 1)/(T\varphi_k)$ cannot be a minimum for large n. The same argument is applicable to the term with $(T - 1)\varphi_{k+1}/T$.

If $\alpha = \alpha_T$, then $\varphi_k = (T + 1)/T$ for all k and, accordingly, we have $\delta_{k,n+1} = \min\{\delta_{k,n}, \delta_{k-1,n}\}$ for all k and for all sufficiently large n. If there is m and k' such that $\delta_{k',m} = 0$, then $\delta_{k',n} = 0$ for all $n \geq m$ and $\delta_{k'+1,m+1} = 0$, which again gives $\delta_{k'+2,m+2} = 0$ and so on. Therefore, we have $\delta_{k,n} = 0$ for all $n > m + T$ and all k. Hence, to complete the proof for this case, we need to elicit a contradiction if $\delta_{k,n}$ is strictly positive for all n and for all k. Since $\delta_{k,n}$ is a nonincreasing sequence of n, we have

$$\delta_{k,n+s} = \min\{\delta_{k,n+s-1}, \delta_{k-1,n+s-1}\} = \min\{\delta_{k,n+s-2}, \delta_{k-1,n+s-2}, \delta_{k-1,n+s-1}\}$$
$$= \min\{\delta_{k,n+s-2}, \delta_{k-1,n+s-1}\} = \min\{\delta_{k,n}, \delta_{k-1,n+s-1}\},$$

for all $s \in \mathbb{N}$. Since $\delta_{k-1,n+s-1}$ should approach zero monotonically as $s \to \infty$, there should be s_0 such that $\delta_{k-1,n+s-1} < \delta_{k,n}$ for all k and for all $s > s_0$. Therefore, we get $\delta_{k,n+s+T} = \delta_{k-1,n+s+T-1} = \delta_{k-2,n+s+T-2} = \delta_{k,n+s}$ for all $s > s_0$. Since δ cannot increase, we conclude that $\delta_{k,n}$ is a constant for all sufficiently large n. If $\delta_{k,n}$ is strictly positive for all n as assumed, C_k cannot be a limit and we arrive at a contradiction. Therefore, there is t_1 such that $c_{t+T} = c_t$ for all $t \geq t_1$ in this case.

If $\alpha \neq \alpha_T$, then there is at least one φ_k such that $(T + 1)/T < \varphi_k$. If ϵ also satisfies $\epsilon/C_k < 1 - (T + 1)/(T\varphi_k)$, then we can write

$$\delta_{k,n+1} = \min\left\{\delta_{k,n}, 1 - \tfrac{T-1}{T}\varphi_{k+1}(1 - \delta_{k+1,n})\right\},$$

for all $n > m_0$. If $\varphi_{k+1} < T/(T - 1)$, then $\delta_{k,n}$ will eventually be smaller than $1 - (T - 1)\varphi_{k+1}/T$ and we have $\delta_{k,n+1} = \delta_{k,n} = 0$ for all large n. On the other hand, if $\varphi_{k+1} = T/(T - 1) > (T + 1)/T$, we have

$$\delta_{k,n+1} = \min\left\{\delta_{k,n}, \delta_{k+1,n}\right\},$$
$$\delta_{k+1,n+1} = \min\left\{\delta_{k+1,n}, 1 - \tfrac{T-1}{T}\varphi_{k+2}(1 - \delta_{k+2,n})\right\}. \tag{33}$$

Since it is impossible for all φ_k to be $T/(T - 1)$, there exists k' such that $\varphi_{k+i} = T/(T - 1)$ for $1 \leq i \leq k'$ and $\varphi_{k+k'+1} < T/(T - 1)$. Therefore,

$$\delta_{k+k',n+1} = \delta_{k+k',n},$$

for all sufficiently large n. Once $\delta_{k+k',m} = 0$, then $\delta_{k+i,n} = 0$ for all $0 \leq i \leq k'$ and for all $n > m + T$ by (33). If $\varphi_{k+k'+1} > (T + 1)/T$, we can repeat the above procedure. If $\varphi_{k+k'+1} = (T + 1)/T$, we have

$$\delta_{k+k'+1,n+1} = \min\left\{\delta_{k+k'+1,n}, \delta_{k+k',n}, 1 - \tfrac{T-1}{T}\varphi_{k+k'+2}(1 - \delta_{k+k'+2,n})\right\} = 0$$

for all sufficiently large n. Hence, the proof is complete. $\qquad\square$

5.3 Non-uniqueness of Periodic Solutions

Proposition 17 and its proof have shown the general periodic solutions of recursion relation (9) for $t > t_1$ to be of the form

$$\chi_t = \chi_{t_1} \prod_{k=t_1+1}^{t} \varphi_k, \quad c_t = c_{t_1} e^{-\nu(t-t_1)} \prod_{k=t_1+1}^{t} \varphi_k,$$

where the φ_k satisfy $\varphi_{T+k} = \varphi_k$ and (31). Since $(T+1)/T \leq e^{\nu} < T/(T-1)$, setting $\varphi_i = e^{\nu}$ for all i satisfies (31), which gives the constant sequence $c_t = c_{t_1}$. We refer to this solution as the homogeneous state. Recall that the homogeneous state is the unique possibility for $\alpha = \alpha_T$, as shown right after (31). By constructing an appropriate sequence (a_n), we now show that any set $\{\varphi_k\}$ that satisfies the conditions (31) can give rise to a periodic solution c_t. Therefore, the periodic solution c_t is not unique and can vary substantially with (a_n) unless $\alpha = \alpha_T$ or $T = 1$.

Proposition 18 *Let*

$$a_t = \max\left\{ \frac{T-i+t-1}{\alpha} \psi_i : 0 \leq i < T \right\}, \quad \psi_i := \prod_{j=1}^{i} \varphi_j, \tag{34}$$

where the φ_j are as in (31) with periodicity $\varphi_{T+j} = \varphi_j$ and we have used the convention $\prod_{j=1}^{0} = 1$. Then

$$\chi_t = \psi_{t+T-1} := \prod_{j=1}^{t+T-1} \varphi_j. \tag{35}$$

Proof To find a_1, we observe that for $0 \leq i < T-1$

$$\frac{T-i-1}{\alpha} \psi_{i+1} = \frac{T-i}{\alpha} \psi_i \frac{T-i-1}{T-i} \varphi_{i+1}$$

$$\leq \frac{T-i}{\alpha} \psi_i \frac{(T-i-1)T}{(T-i)(T-1)} = \frac{T-i}{\alpha} \psi_i \frac{T^2-(i+1)T}{T^2-(i+1)+i} \leq \frac{T-i}{\alpha} \psi_i,$$

where we have used $\varphi_i \leq T/(T-1)$. Therefore, we get

$$\chi_1 = a_1 = \frac{T}{\alpha} = \prod_{j=1}^{T} \varphi_j = \psi_T,$$

which is (35) for $t = 1$. Note that this χ_1 is trivially valid for $T = 1$.

Now assume (35) is valid up to $t = n$. Then,

$$\chi_{n+1} = \max\left\{ \frac{T-i+n}{\alpha} \psi_i : 0 \leq i \leq n+T-1 \right\}.$$

For $i \leq n$, we have

$$\frac{T-i+n}{\alpha} \psi_i = \frac{T-i+n+1}{\alpha} \psi_{i-1} \frac{T-i+n}{T-i+n+1} \varphi_i$$

$$\geq \frac{T-i+n+1}{\alpha} \psi_{i-1} \frac{(T-i+n)(T+1)}{(T-i+n+1)T}$$

$$= \frac{T - i + n + 1}{\alpha} \psi_{i-1} \frac{T^2 + (n + 1 - i)T + n - i}{T^2 + (n + 1 - i)T}$$

$$\geq \frac{T - i + n + 1}{\alpha} \psi_{i-1},$$

and for $n + T - 1 > i \geq n$ and $T > 1$, we have

$$\frac{T - i + n}{\alpha} \psi_i = \frac{T - i + n - 1}{\alpha} \psi_{i+1} \frac{T - i + n}{T - i + n - 1} \frac{1}{\varphi_{i+1}}$$

$$\geq \frac{T - i + n - 1}{\alpha} \psi_{i+1} \frac{(T - i + n)(T - 1)}{(T - i + n - 1)T}$$

$$= \frac{T - i + n - 1}{\alpha} \psi_{i+1} \frac{T^2 + (n - 1 - i)T + i - n}{T^2 + (n - 1 - i)T}$$

$$\geq \frac{T - i + n - 1}{\alpha} \psi_{i+1}.$$

Therefore, we have

$$\chi_{n+1} = \frac{T}{\alpha} \psi_n = \psi_n \prod_{j=n+1}^{T+n} \varphi_j = \prod_{j=1}^{T+n} \varphi_j.$$

Induction completes the proof. □

Now we illustrate that any allowed set of φ_j's can appear in the actual branching process by choosing an appropriate initial condition. For a realization of (34) in the branching process, consider an initial condition such that there are T different mutant classes with fitness $f_i := f^{\psi_i/\alpha}$ ($0 \leq i < T$) and the number N_i of individuals with fitness f_i is

$$N_i = \left\lfloor f^{(T-i-1)\psi_i/\alpha} \right\rfloor.$$

Notice that this initial condition with $\varphi_j = e^{\nu'}$ together with a shift in time was used in the proof of Lemma 16. In the limit $f \to \infty$ as in Sect. 3, $X(t)$ is well approximated by f^{χ_t} with a_t in (34).

In the above discussion, we have illustrated that any permissible set of φ_j's can be realized by choosing an appropriate initial condition. Now we argue that a surviving outcome with an arbitrary initial condition should approach such a permissible set, but which values of the φ_j's are realized may depend on the stochasticity in the early time regime. In the original branching process, the sequence (a_n) depends both on the initial condition and the stochastic evolution in the early time regime before the deterministic approximation through the recursion relation (9) becomes valid. To see this, we recall from Sect. 3 how the recursion relation arises from the stochastic process. Since on survival the total population size as well as the largest fitness increases indefinitely, there should be a generation t_0 such that $X(t_0) > K$ for any preassigned K. Let W_0 be the largest fitness at generation t_0, define $Y = X(t_0)$ and introduce a shifted time variable $t' = t - t_0$ with $\tilde{X}(t') := X(t' + t_0)$. If K is extremely large, $\tilde{X}(t')$ can be well approximated as $\tilde{X}(0) = Y$,

$$\tilde{X}(1) = Y^{a_1} + 1 \approx Y^{\chi_1}, \quad W_1 \approx Y^{\chi_1/\alpha},$$

$$\tilde{X}(2) = Y^{a_2} + Y^{\chi_1/\alpha} + 1 \approx Y^{\chi_2}, \quad W_2 \approx Y^{\chi_2/\alpha},$$

$$\tilde{X}(3) = Y^{a_3} + Y^{2\chi_1/\alpha} + Y^{\chi_2/\alpha} + 1 \approx Y^{\chi_3}, \quad W_3 \approx Y^{\chi_3/\alpha}$$
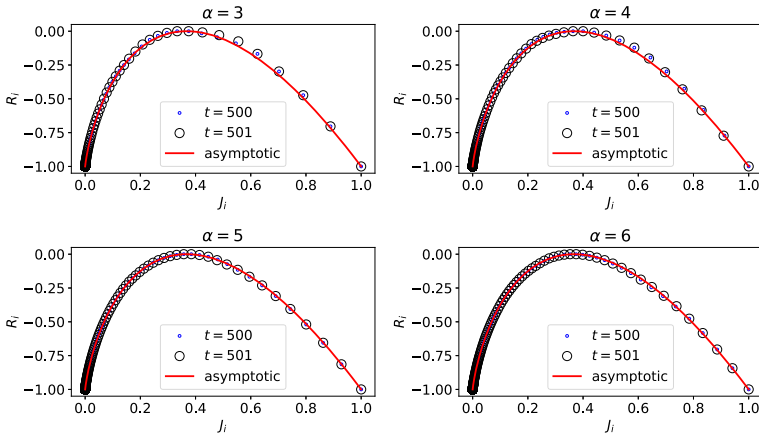
Fig. 4 Plot of R_i vs J_i at $t = 500$ and 501 for $\alpha = 3, 4, 5, 6$. For comparison, the asymptotic prediction (36) is depicted as a solid curve. For $\alpha = 3$ and 4, the changes of the empirical fitness distributions between generations 500 and 501 are still visible, but the distributions become indistinguishable from the asymptotic form (36) for $\alpha \geq 5$

where Y^{a_n} is the population size of all mutant classes that appeared prior to generation t_0. Since $Y^{a_n} \leq Y W_0^n$, we naturally have $\lim_{n \to \infty} a_n e^{-\nu n} = 0$, and (a_n) is a permissible sequence that can be entered into the recursion relation (9).

5.4 Empirical Fitness Distribution for Large α

Whereas the preceding subsection has shown that the empirical fitness distribution at long times is generally non-universal, we will now argue that it nevertheless has a well-defined limit for $\alpha \to \infty$. Let us begin with the homogeneous state. In this case,

$$J_i(t) = e^{-\nu(t-i)}, \quad P(t) \equiv \alpha(e^\nu - 1),$$

$$R_i(t) = \frac{(t-i)}{\alpha} J_i(t) - 1 = -\frac{1}{\nu\alpha} J_i(t) \log J_i(t) - 1.$$

Since $\nu\alpha \to 1/e$ as $\alpha \to \infty$, the homogeneous state for all sufficiently large α is well described by

$$R_i \approx -e J_i \log J_i - 1, \tag{36}$$

and the mean log fitness converges to $P = \frac{1}{e}$. Moreover, since

$$\frac{T}{T-1} - \frac{T+1}{T} = \frac{1}{(T-1)T} = O(T^{-2})$$

and $T/\alpha \to e$ as $\alpha \to \infty$, in this limit all periodic solutions that satisfy the constraints (31) become close to homogeneous, $\varphi_i = e^\nu + O(\alpha^{-2})$. Therefore, we conjecture that the empirical fitness distribution on survival has (36) as a limit distribution for $\alpha \to \infty$. As an illustration, in Fig. 4 we compare (36) to numerical solutions of the recursion relation for $\alpha = 3, 4, 5, 6$. The numerical data are hardly distinguishable from (36) already for $\alpha = 5$.

6 Summary and Discussion

In this article we have provided a detailed characterization of the superexponential population growth in two closely related stochastic models of evolution. To the best of our knowledge, this is the first rigorous analysis of a branching process with selection and mutations where the random fitness values (rather than the fitness differences [22]) are drawn from an unbounded probability distribution. A remarkable feature of the models considered here is the emergence of an integer-valued time scale T which depends (discontinuously) on the index α of the underlying Fréchet distribution. As a consequence, the empirical fitness distribution displays oscillations with period T, a phenomenon that has been observed previously in certain models that include sexual reproduction [38, 39]. A partial understanding of the periodic behaviour of the population structure was achieved in a deterministic approximation. Further work on this problem is needed, addressing in particular how the stochastic initial phase of the process determines the non-universal aspects of the asymptotic population distribution.

It is instructive to compare our findings for the branching process to the earlier analysis of a stochastic fixed finite population version of Kingman's model in [13]. In both cases the long-time behaviour is dominated by, and can quantitatively understood in terms of extremal mutation events in the past. However, in the fixed finite population model the likelihood of generating mutants that exceed the current population fitness declines with time, and the dynamics reduces to a modified record process, where the takeover of the population by a fit mutant is instantaneous compared to the waiting time for the next fitter mutant. As a consequence, the population at time t is dominated by a mutant that arose at a time of order t in the past. By contrast, in the branching process with Fréchet-type distributions, the declining probability of exceeding the current fitness is compensated by the rapid growth of the population in such a way that the time lag since the birth of the currently dominant mutant takes on a fixed value T. Moreover, the branching process never enters the regime of rare sequential fixation events associated with the decreasing supply of beneficial mutations in the finite population setting. Instead, the population attains a nontrivial stationary clonal structure which is approximately described in Sect. 5.

It is reasonable to expect that the growth of the population fitness in the branching process is intermediate between that of the fixed finite population model [13] and the deterministic infinite population model [5]. For Fréchet type fitness distributions the deterministic model is ill-defined, but the analysis of the fixed finite population model predicts a polynomial increase of the fitness with exponent $1/\alpha$ [13], which is indeed much slower than the superexponential growth in the branching process. For unbounded Gumbel type distributions the growth law of the fitness is known for infinite as well as for finite populations [5, 13]. The corresponding behaviour of the branching process will be addressed in future work.

Data Availability All material needed to reproduce the work is fully contained in the paper.

References

1. Park, S.-C., Simon, D., Krug, J.: The speed of evolution in large asexual populations. J. Stat. Phys. **138**, 381–410 (2010)
2. Wiser, M.J., Ribeck, N., Lenski, R.E.: Long-term dynamics of adaptation in asexual populations. Science **342**, 1364–1367 (2013)
3. Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., Desai, M.M.: The dynamics of molecular evolution over 60,000 generations. Nature **551**, 45–50 (2017)
4. Bürger, R.: The Mathematical Theory of Selection, Recombination, and Mutation. Wiley, Chichester (2000)
5. Kingman, J.F.C.: A simple model for the balance between selection and mutation. J. Appl. Probab. **15**, 1–12 (1978)
6. Waxman, D., Peck, J.R.: Pleiotropy and the preservation of perfection. Science **279**, 1210–1213 (1998)
7. Dereich, S., Mörters, P.: Emergence of condensation in Kingman's model of selection and mutation. Acta Appl. Math. **127**, 17–26 (2013)
8. Yuan, L.: A generalization of Kingman's model of selection and mutation and the Lenski experiment. Math. Biosci. **285**, 61–67 (2017)
9. Yuan, L.: Kingman's model with random mutation probabilities: convergence and condensation II. J. Stat. Phys. **181**, 870–896 (2020)
10. Yuan, L.: Kingman's model with random mutation probabilities: convergence and condensation I. Adv. Appl. Probab. **54**, 311–335 (2022)
11. Hwang, S., Schmiegelt, B., Ferretti, L., Krug, J.: Universality classes of interaction structures for NK fitness landscapes. J. Stat. Phys. **172**, 226–278 (2018)
12. Kimura, M., Crow, J.F.: The number of alleles that can be maintained in a finite population. Genetics **49**, 725–738 (1964)
13. Park, S.-C., Krug, J.: Evolution in random fitness landscapes: the infinite sites model. J. Stat. Mech. Exp. **2008**(4), 04014 (2008)
14. Sibani, P., Brandt, M., Alstrøm, P.: Evolution and extinction dynamics in rugged fitness landscapes. Int. J. Mod. Phys. B **12**, 361–391 (1998)
15. Bianconi, G., Fichera, D., Franz, S., Peliti, L.: Modeling microevolution in a changing environment: the evolving quasispecies and the diluted champion process. J. Stat. Mech. **2011**, 08022 (2011)
16. Desai, M.M., Fisher, D.S.: Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176**, 1759–1798 (2007)
17. Park, S.-C., Krug, J.: Clonal interference in large populations. Proc. Natl. Acad. Sci. USA **104**, 18135–18140 (2007)
18. Yu, F., Etheridge, A., Cuthbertson, C.: Asymptotic behavior of the rate of adaptation. Ann. Appl. Probab. **20**, 978–1004 (2010)
19. Fisher, D.S.: Asexual evolution waves: fluctuations and universality. J. Stat. Mech. **2013**, 01011 (2013)
20. Kelly, M.: Upper bound on the rate of adaptation in an asexual population. Ann. Appl. Probab. **23**, 1377–1408 (2013)
21. Schweinsberg, J.: Rigorous results for a population model with selection I: evolution of the fitness distribution. Electron. J. Probab. **22**, 1–94 (2017)
22. Durrett, R., Foo, J., Leder, K., Mayberry, J., Michor, F.: Evolutionary dynamics of tumor progression with random fitness values. Theor. Popul. Biol. **78**, 54–66 (2010)
23. Durrett, R.: Population genetics of neutral mutations in exponentially growing cancer cell populations. Ann. Appl. Probab. **23**, 230–250 (2013)
24. Angaji, A., Velling, C., Berg, J.: Stochastic clonal dynamics and genetic turnover in exponentially growing populations. J. Stat. Mech. **2021**, 103502 (2021)
25. Dereich, S., Mailler, C., Mörters, P.: Nonextensive condensation in reinforced branching processes. Ann. Appl. Probab. **27**, 2539–2568 (2017)

26. Mailler, C., Mörters, P., Senkevich, A.: Competing growth processes with random growth rates and random birth times. Stoch. Proc. Appl. **135**, 183–226 (2021)

27. Resnick, S.: Extreme Values, Regular Variation, and Point Processes. Springer, New York (1987)

28. Joyce, P., Rokyta, D.R., Beisel, C.J., Orr, H.A.: A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. Genetics **180**, 1627–1643 (2008)

29. Orr, H.A.: The population genetics of beneficial mutations. Philos. Trans. R. Soc. B **365**, 1195–1201 (2010)

30. Bataillon, T., Bailey, S.F.: Effects of new mutations on fitness: insights from models and data. Ann. N. Y. Acad. Sci. **1320**, 76–92 (2014)

31. Das, S.G., Krug, J.: Unpredictable repeatability in molecular evolution. Proc. Natl. Acad. Sci. USA **119**, 2209373119 (2022)

32. Schenk, M.F., Szendro, I.G., Krug, J., de Visser, J.A.G.M.: Quantifying the adaptive potential of an antibiotic resistance enzyme. PLoS Genet. **8**, 1002783 (2012)

33. Bank, C., Hietpas, R.T., Wong, A., Bolon, D.N., Jensen, J.D.: A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. Genetics **175**, 841–852 (2014)

34. Foll, M., Poh, Y.-P., Renzette, N., Ferrer-Admetlla, A., Bank, C., Shim, H., Malaspinas, A.-S., Ewing, G., Liu, P., Wegmann, D., Caffrey, D.R., Zeldovich, K.B., Bolon, D.N., Wang, J.P., Kowalik, T.F., Schiffer, C.A., Finberg, R.W., Jensen, J.D.: Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet. **10**, 1004185 (2014)

35. Tokutomi, N., Nakai, K., Sugano, S.: Extreme value theory as a framework for understanding mutation frequency distribution in cancer genomes. PLoS ONE **16**, 0243595 (2021)

36. Davies, P.L.: The simple branching process: a note on convergence when the mean is infinite. J. Appl. Probab. **15**, 466–480 (1978)

37. Harris, T.E.: The Theory of Branching Processes. Springer, Berlin (1963)

38. Park, S.-C., Krug, J.: Rate of adaptation in sexuals and asexuals: a solvable model of the Fisher–Muller effect. Genetics **195**, 941–955 (2013)

39. Pearce, M.T., Fisher, D.S.: Rapid adaptation in large populations with very rare sex: scalings and spontaneous oscillations. Theor. Popul. Biol. **129**, 18–40 (2019)