Branching with selection and mutation II: Mutant fitness of Gumbel type

Su-Chan Park, Joachim Krug, Peter Mörters

Abstract

We study a model of a branching process subject to *selection*, modeled by giving each family an individual fitness acting as a branching rate, and *mutation*, modeled by resampling the fitness of a proportion of offspring in each generation. For two large classes of fitness distributions of Gumbel type we determine the growth of the population, almost surely on survival. We then study the empirical fitness distribution in a simplified model, which is numerically indistinguishable from the original model, and show the emergence of a Gaussian travelling wave.

1 Introduction

We consider the branching processes with selection and mutation introduced in [1]. These are models of a population evolving in discrete non-overlapping generations with model parameters given by a probability distribution μ on $(0,\infty)$, which serves as a means to sample a random fitness of a mutant, and a mutation probability $\beta \in (0,1)$. For later reference, we denote the tail function by $G(x) := \mu((x,\infty))$. Note that G is a right-continuous-left-limit function that may be discontinuous.

A brief description of the two model variants goes as follows: In each generation a population consists of finitely many individuals each equipped with a positive fitness. Any individual lives only for one generation. Every generation produces a random number of offspring, which is Poisson distributed with the mean given by the sum over all the fitnesses of the individuals in the generation. Now every offspring individual independently

- with probability $1-\beta$ randomly *selects* a parent with a probability proportional to its fitness. The offspring becomes an individual of the next generation with the fitness inherited from the parent;
- otherwise, with probability β, it is a mutant and gets a fitness randomly sampled from μ.
  - In the *fittest mutant model (FMM)* only one mutant with largest fitness among all mutants, if it exists, joins the next generation and the others die immediately.
  - In the *multiple mutant model (MMM)* all mutants join the next generation.

We write $X(t)$ for the number of individuals in generation t, irrespective of what initial condition is used and which model variant is under consideration. Further discussion of the motivation behind this model can be found in the first paper of this series [1]. Other branching models including selection or mutation are [2, 3, 4] or [5]. Similar models have been applied for the description of the genetic structure of proliferating tumors and growing populations of pathogens [6, 7, 8]

Our focus in this paper is on the case of unbounded fitness distributions μ with light tails at infinity, but to put this into context we briefly review known results first on bounded and second on unbounded heavy-tailed random variables.

Suppose first that $a :=$esssup $\mu < \infty$ let $\lambda^* := (1-\beta)a$. In the MMM, if $\beta \int \frac{a}{a-x}\,\mu(dx) \geq 1$ there is a unique $\lambda \geq \lambda^*$ such that

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

Then, almost surely on survival, we have

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

Otherwise, and always in the FMM, we have, almost surely on survival,

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

This is shown in [9] for a continuous-time variant of the model and the proof extends to the MMM. For the FMM note that in generation t there are at most $t+1$ families with fitness $W_0, \ldots, W_t$ present, each growing at rate $\log((1-\beta)W_i)$. The overall growth rate is therefore bounded from above by $\log \lambda^*$ and also from below as $\limsup W_t = a$ almost surely on survival. So, irrespective of the finer details of μ, we see exponential growth of the population.

In the case of a slowly decreasing tail at infinity, i.e. when the tail function G is regularly varying with index $-\alpha$, for some $\alpha > 0$, we have doubly exponential growth. We show in [1] that, for T the unique integer such that

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

in either MMM or FMM, almost surely on survival,

$$\lim_{t\to\infty} \frac{\log\log X(t)}{t} = \frac{1}{T} \log \frac{T}{\alpha},$$

i.e. we have doubly exponential growth of the population. The present paper is concerned with unbounded fitness distributions with *light tail at infinity*. In analogy to the classification of distribution as extremal types we denote this class of fitness distributions as Gumbel type [10]. The classification of fitness distributions in terms of extreme value classes plays an important role in the theory of evolutionary adaptation [11]. In this context it has been argued that the Gumbel type is the most relevant case biologically [12, 13, 14].

For unbounded fitness distributions of Gumbel type the population grows at a rate between exponential and doubly exponential. This is a wide range that cannot be easily covered by a single functional expression. Therefore we introduce parametrised subclasses of fitness distributions and show how the population grows for these subclasses in dependence of the parameters. Before stating our full results in Section 2 we describe an interesting example to give a flavour.

We look at fitness distributions with stretched exponential tail satisfying

$$\lim_{x\to\infty} \frac{\log(1/G(x))}{x^\alpha L(x)} = 1,$$

for a slowly varying function L and $\alpha > 0$. In this case, for both MMM and FMM we show in Theorem 1 that the population grows like

$$\lim_{t\to\infty} \frac{\log X(t)}{t\log t} = \frac{1}{\alpha}, \qquad \text{almost surely on survival.}$$

The superexponential growth is driven by the fitness W_t of the fittest mutant in generation t satisfying

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

In Section 3 we describe the subtle interplay of population size and fittest mutant heuristically in terms of a differential equation. Simulations demonstrated in Section 7 show that the distribution of fitness in a positive proportion of the population in generation t concentrates around the value

$$v(t) := \alpha^{-1/\alpha} t^{1/\alpha} L(t^{1/\alpha})$$

in the shape of a Gaussian travelling wave of width $v(t)/\sqrt{\alpha t}$. In Theorem 4 we prove this phenomenon rigorously for a simplified model where the driving fitness W_t is replaced by its deterministic asymptotics.

The rest of this paper is organised as follows. Full results on the growth of the population and the driving fitness are formulated as Theorem 1 and 2 in Section 2. The section also formulates, as Theorem 3, the conjectured behaviour of the travelling wave for the full model. Section 3 heuristically describes the interplay of these quantities. Section 4 contains preparation for the proofs of Theorem 1, given in Section 5, and Theorem 2, given in Section 6. Section 7 explains the approximations needed to simulate and prove the travelling wave result restated now in rigorous form as Theorem 4. We finish the paper with concluding remarks in Section 8.

2 Main results

In the FMM, $X(t)$ is generally different from the total number of offspring of all particles in generation $t-1$. We therefore denote by $\Xi(t)$ the total number of offspring of all individuals in generation $t-1$, including immediately dead ones, if there are any. By Q_t we denote the largest fitness *in the population* in generation $t \geq 0$ and by W_t the largest fitness among all mutants in generation $t \geq 1$. Note that $W_t \leq Q_t$ and W_t can be strictly smaller than Q_t. The number of non-mutated descendants in generation $s \geq t$ of the fittest mutant in generation t will be denoted by $N_t(s)$ with the convention that $N_t(t) = 1$. For convenience we set $N_t(s) = 0, W_t = 0$ if there is no mutant in generation t and $Q_t = 0$ if $X(t) = 0$. Also set $\Xi(0) = X(0)$, $W_0 = Q_0$.

2.1 Tail functions

To classify the decay of the tail function G in a way that allows the description of the growth rates of the population size, we denote by $\log^{(n)}$ the nth iterated logarithm, write $f_1(t) \sim f_2(t)$ to mean that the ratio of the two expressions converges to one as t goes to infinity, and assume

$$\log^{(n_1)}(1/G(x)) \sim \big(\log^{(n_2)}(x)\big)^{\tilde{\alpha}} L\big(\log^{(n_2)}(x)\big), \tag{1}$$

where n_1, n_2 are non-negative integers, α is a positive number, and $L(x)$ is assumed to satisfy[1]

$$\lim_{x\to\infty} \frac{L(x)}{x^\varepsilon} = \lim_{x\to\infty} \frac{1}{L(x)x^\varepsilon} = 0, \tag{2}$$

[1]If L is a slowly varying function, then this condition is naturally satisfied.

for any $\varepsilon > 0$. Apart from this assumption, henceforth called **(A1)**, we use three further technical assumptions on L in (1), namely

(A2) If a positive function ℓ satisfies (2), then $\lim_{x\to\infty} \frac{L(x\ell(x))}{L(x)} = 1$.

(A3) L is four-times continuously differentiable, at least for sufficiently large argument.

(A4) $\lim_{x\to\infty} \left(\frac{d}{d\log x}\right)^j \log L(x^\gamma) = 0$, for nonnegative integer j and positive real γ.

Assumption **(A2)** will be used in Section 5. It is a stronger condition than L being a slowly varying function. Assumption **(A3)** will be used in Section 7. Note that even if G is discontinuous, we can, in most cases, find a four-times continuously differentiable L. Assumption **(A4)** will be used in the proof of Lemma 5.2 and in Section 7.2.

As an example of L satisfying all four assumptions, we consider

$$L(x) = \prod_{k=1}^{m} (\log^{(k)}(x))^{\gamma_k} \tag{3}$$

with real γ_k's. Obviously, (3) cannot exhaust all functions satisfying the above four assumptions; an example that does not take the form (3) is $\exp(\sqrt{\log x})$. The proofs of the main theorems apply to any function L that satisfies the above four assumptions.

In this paper, we are interested in Gumbel type tail functions with unbounded support, meaning that at infinity G decays faster than polynomially, i.e., for any positive γ,

$$\lim_{x\to\infty} x^\gamma G(x) = 0. \tag{4}$$

We now figure out[2] for which parameters n_1, n_2 and $\tilde{\alpha}$, (4) holds. If $n_1 < n_2$, then G satisfies

$$\lim_{x\to\infty} \frac{x^{-\varepsilon}}{G(x)} = 0 \tag{5}$$

for any positive ε. As this G decays slower than any Fréchet type tail function, the long-time evolution is dominated by the largest fitness alone as in the Fréchet type with $\alpha < 0.5$, as studied in [1]. If $n_1 > n_2$, then G satisfies (4), which will be our concern. We define the n-th iterated exponential function $\exp^{(n)}$ as the inverse of $\log^{(n)}$ with the convention $\exp^{(0)}(x) = \log^{(0)}(x) = x$. In case $n_2 > 0$, we have a rough bound for sufficiently large x as

$$\begin{aligned}
1/G(x) &\geq \exp^{(n_1)}\left(\log^{(n_2)}(x)^{\tilde{\alpha}-\varepsilon}\right) \\
&= \exp^{(n_1-n_2)}\left(\exp^{(n_2)}\left(\log^{(n_2)}(x)^{\tilde{\alpha}-\varepsilon}\right)\right) \\
&= \exp^{(n_1-n_2)}\left(\exp^{(n_2+1)}\left((\tilde{\alpha}-\varepsilon)\log^{(n_2+1)}(x)\right)\right) \geq \exp^{(n_1-n_2)}\left(x^{\tilde{\alpha}-\varepsilon}\right),
\end{aligned}$$

where we have used Lemma 5.1 for the last inequality, and

$$1/G(x) \leq \exp^{(n_1)}\left(\log^{(n_2-1)}(x)\right) = \exp^{(n_1-n_2+1)}(x). \tag{6}$$

[2]It is of course possible that $G(x)$ satisfies (4) but not (1).

In this context, limiting ourselves to the case with $n_1 > n_2 = 0$ would give a guide for $n_1 > n_2 > 0$. For example, inspecting Theorem 1 suggests that almost surely on survival

$$\lim_{t\to\infty} \frac{\log^{(2)}(X(t))}{\log t} = 1$$

for any case with $n_1 > n_2 \geq 0$. The remaining case is $n_1 = n_2$. If $n_1 = n_2 = 0$, then G does not satisfy (4). In fact, this G becomes a Fréchet-type tail function already studied in [1]. If $n_1 = n_2 > 0$, then how fast G decays is determined by $\tilde{\alpha}$. If $0 < \tilde{\alpha} < 1$, then G satisfies (5). If $\tilde{\alpha} > 1$, then G satisfies (4). If $\tilde{\alpha} = 1$, then how fast G decays depends on the explicit form of L. For example, assume $L(x) = (\log x)^\gamma \bar{L}(\log x)$ with $\bar{L}$ to satisfy (2). If $\gamma > 0$, then G satisfies (4), while if $\gamma < 0$, then G satisfies (5). If $\gamma = 0$, then how fast G decays depends on the explicit form of $\bar{L}$. In this sense, it is difficult, if not impossible, to write all possible tail functions that satisfy (4). We take a rather special form of L for $\tilde{\alpha} \geq 1$; see (8). We only study the case $n_1 = n_2 = 1$, but the case with $n_1 = n_2 > 1$ can be easily studied using the techniques developed in this paper.

In this paper, we therefore limit ourselves to two cases. The first case that corresponds to $n_2 = 0$ and $n_1 = n \geq 1$ with $\tilde{\alpha} = \alpha > 0$ is

$$\log^{(n)}(1/G(x)) \sim g_{\mathrm{I}}(x) := x^\alpha L(x). \tag{7}$$

The second case that corresponds to $n_1 = n_2 = 1$ with $\tilde{\alpha} \geq 1$ is

$$\frac{\log(1/G(x))}{\log x} \sim g_{\mathrm{II}}(x) := g_{\mathrm{I}}(\log^{(n)}(x)), \tag{8}$$

where $n \geq 1$ and $\alpha > 0$. Note that for the second case $\tilde{\alpha} = 1 + \alpha > 1$ for $n = 1$ and $\tilde{\alpha} = 1$ for $n \geq 2$. From now on, n and α are reserved for this role, with n called the *tail index* and α the *tail parameter*. When G satisfies (7), we will say that G is of type I and when G satisfies (8), we will say that G is of type II. Note that not only do the two types of decay not cover the entire Gumbel class, but conversely (4) alone cannot guarantee that G falls into the Gumbel class. For instance, consider $\log G(x) = -x - \sin(x)$, which is of type I but does not belong to the Gumbel class (see, e.g., [10]).

2.2 Statement of theorems

Our main concern is how $X(t)$, W_t, and the empirical fitness distribution (EFD) behave at large times t on survival. The EFD is defined via its cumulative distribution function $\Psi(f,t)$ as

$$\Psi(f,t) := \frac{1}{X(t)} \sum_{i=1}^{X(t)} \Theta(f - F_i), \tag{9}$$

where F_i is the fitness of i-th individual and $\Theta(x)$ is the Heaviside step function with $\Theta(0) = 1$. We denote the mean and the standard deviation of $\Psi(f,t)$ by S_t and σ_t, respectively. In case that no individual is left at t, we define $\Psi(f,t) = 1$ for $f \geq 0$ and $S_t = \sigma_t = 0$. We define the survival event $\mathfrak{A}$ and survival probability p_s as

$$\mathfrak{A} := \{X(t) \neq 0 \text{ for all } t\}, \quad p_s := \mathbb{P}(\mathfrak{A}).$$

Needless to say, p_s depends on the initial condition, but the initial condition dependence does not play any role in what follows. Now we state the main theorems.

Theorem 1. *If G is of type I, then almost surely on survival*

$$\lim_{t\to\infty}\frac{\log X(t)}{t\log^{(n)}(t)}=\frac{1}{\alpha},\quad \lim_{t\to\infty}\frac{W_t}{u_n(t)}=1, \tag{10}$$

where

$$u_n(t):=\left(\log^{(n-1)}(t)\right)^{1/\alpha}\omega_W\left(\log^{(n-1)}(t)\right), \tag{11}$$

$$\omega_W(y):=\left(\frac{\log y}{\alpha}\right)^{\delta_{n,1}/\alpha}\left[L\left(y^{1/\alpha}\right)\right]^{-1/\alpha}, \tag{12}$$

with $\delta_{n,1}$ to be the Kronecker delta symbol.

Theorem 2. *If G is of type II, then almost surely on survival*

$$\lim_{t\to\infty}\frac{\log^{(2)}(X(t))}{\log t}=1+\frac{1}{\alpha},\quad \lim_{t\to\infty}\frac{\log^{(2)}(W_t)}{\log t}=\frac{1}{\alpha},$$

for $n=1$,

$$\lim_{t\to\infty}\frac{\log^{(3)}(X(t))}{\log t}=\lim_{t\to\infty}\frac{\log^{(3)}(W_t)}{\log t}=\frac{1}{1+\alpha},$$

for $n=2$, and

$$\lim_{t\to\infty}\frac{1}{\log^{(n-1)}(t)}\log\left(\frac{\log^{(2)}(X(t))}{t}\right)=\lim_{t\to\infty}\frac{1}{\log^{(n-1)}(t)}\log\left(\frac{\log^{(2)}(W_t)}{t}\right)=-\alpha,$$

for $n\geq 3$.

Based on simulations, we conjecture the following theorem regarding the EFD formulated in the case of the FMM. A rigorously proved version of this result and more details on our simulations will be given in Section 7.

Theorem 3 (Conjecture). *For each type I tail function, there are positive functions $v(t)$ and $\mathfrak{s}(t)$ such that*

$$\lim_{t\to\infty}v(t)=\infty,\quad \lim_{t\to\infty}\frac{\mathfrak{s}(t)}{v(t)}=0,$$

and almost surely on survival

$$\lim_{t\to\infty}\Psi(v(t)+y\mathfrak{s}(t),t)=\Upsilon(y),$$

where

$$\Upsilon(y):=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}\exp\left(-\frac{1}{2}x^2\right)dx. \tag{13}$$

In particular, if $n=1$ and $\alpha>2$ or if $n\geq 2$, then $\mathfrak{s}(t)\to 0$ as $t\to\infty$ and, for $y\neq 0$,

$$\lim_{t\to\infty}\Psi(v(t)+y,t)=\Theta(y)\qquad \textit{almost surely on survival.}$$

Remark 2.1. *A similar statement is conjectured for the MMM where a fraction $1-\beta$ of the mass in the EFD enters the travelling wave and a fraction β remains in the bulk.*

Corollary 2.1. *Given Theorem 3 the empirical mean fitness satisfies*

$$\lim_{t\to\infty} \frac{S_t}{v(t)} = 1 \qquad \textit{almost surely on survival.}$$

Proof. Fix $\varepsilon > 0$. By Markov's inequality, we have $1 - \Psi(\frac{v(t)}{1+\varepsilon}, t) \le (1+\varepsilon)\frac{S_t}{v(t)}$. As $\mathfrak{s}(t)/v(t) \to 0$ as $t \to \infty$, Theorem 3 implies that $\lim_{t\to\infty} \Psi(\frac{v(t)}{1+\varepsilon}, t) = 0$. Therefore, almost surely on survival, we have $\liminf_{t\to\infty} \frac{S_t}{v(t)} \ge \frac{1}{1+\varepsilon}$. As ε was arbitrary we have, almost surely on survival, $\liminf_{t\to\infty} \frac{S_t}{v(t)} \ge 1$.

Now assume $\limsup_{t\to\infty} \frac{S_t}{v(t)} > 1$. Then there is $\varepsilon' > 0$ and a strictly increasing sequence $(t_k)_{k=1}^{\infty}$ such that $S_{t_k} \ge v(t_k)(1+2\varepsilon')$ for all k. As Theorem 3 implies $\lim_{t\to\infty} \Psi(\frac{1+2\varepsilon'}{1+\varepsilon'} v(t), t) = 1$, we have $\lim_{k\to\infty} \Psi(\frac{S_{t_k}}{1+\varepsilon'}, t_k) = 1$, which contradicts to the definition of S_t. Therefore, we conclude that almost surely on survival $\limsup_{t\to\infty} \frac{S_t}{v(t)} \le 1$, which along with the lower bound gives $S_t \sim v(t)$. $\square$

In Section 7, we will modify our model so that a version of Theorem 3 can be proved.

3 Heuristic guide to Theorems 1 and 2

Before delving into the proofs, we first sketch the idea behind Theorems 1 and 2 by a mean-field type analysis of the MMM for a strictly decreasing continuous G with $g_{\mathrm{I}}(x) \sim x^{\alpha}$. Let us assume that at certain time t, the population size $X(t)$ is very large. Once $X(t)$ is given, W_t is sampled as $Z = [1 - \beta G(W_t)]^{X(t)}$, where Z is uniformly distributed on $(0, 1)$; see Lemma 4.8. Neglecting fluctuation in the sense that $-\log G(W_t) \approx \log X(t)$, we have

$$\log W_t \approx \frac{1}{\alpha} \log^{(n+1)}(X(t)) \tag{14}$$

for type I and

$$\log W_t \approx \begin{cases} [\log X(t)]^{1/(1+\alpha)}, & n = 1, \\ \left[\log^{(n)}(X(t))\right]^{-\alpha} \log X(t), & n \ge 2, \end{cases} \tag{15}$$

for type II. Since the mean fitness S_t is anticipated not to be larger than W_t and $\log X(t+1) \approx \log X(t) + \log S_t$, we have $\log X(t+1) - \log X(t) \le \log W_t$. Treating t as a continuous variable and setting $y = \log X(t)$, we assume that the solutions of the differential equations

$$\frac{dy}{dt} = \frac{1}{\alpha} \log^{(n)}(y) \tag{16}$$

for type I and

$$\frac{dy}{dt} = \begin{cases} y^{1/(1+\alpha)}, & n = 1, \\ y/(\log^{(n-1)}(y))^{\alpha}, & n \ge 2, \end{cases} \tag{17}$$

for type II give the upper bound for the corresponding $\log X(t)$. The asymptotic behaviour of the solution of (16) can be found as

$$\frac{t}{\alpha} = \int^y \frac{dx}{\log^{(n)}(x)} = \frac{y}{\log^{(n)}(y)} + \int^y \frac{1}{(\log^{(n)}(x))^2} \left(\prod_{k=1}^{n-1} \frac{1}{\log^{(k)}(x)} \right) dx \approx \frac{y}{\log^{(n)}(y)},$$

which gives

$$y \approx \frac{t}{\alpha} \log^{(n)}(y) \approx \frac{t}{\alpha} \log^{(n)}(t),$$

where we have used **(A2)** for $L(x) = \log^{(n)}(x)$. In a similar manner, we find the asymptotic solution of (17) as $y \approx t^{1+1/\alpha}$ if $n = 1$, $y \approx \exp(t^{1/(1+\alpha)})$ if $n = 2$ and $y \approx \exp(t(\log^{(n-2)}(t))^{-\alpha})$ if $n \geq 3$. Accordingly, we anticipate

$$\log X(t) \lessapprox \frac{1}{\alpha} t \log^{(n)}(t)$$

for type I and

$$\log X(t) \lessapprox \begin{cases} t^{1+1/\alpha}, & n = 1, \\ \exp\left(t^{1/(1+\alpha)}\right), & n = 2, \\ \exp\left(t\left(\log^{(n-2)}(t)\right)^{-\alpha}\right), & n \geq 3. \end{cases}$$

for type II.

Theorems 1 and 2 actually state that to treat the above inequalities as equalities gives a good approximation. If the inequalities are indeed equalities, then we expect

$$W_t \approx \left(\log^{(n-1)}(t \log^{(n)}(t))\right)^{1/\alpha}$$

for type I and

$$W_t \approx \begin{cases} \exp\left(t^{1/\alpha}\right), & n = 1, \\ \exp\left(t^{-\alpha/(1+\alpha)} \exp\left(t^{1/(1+\alpha)}\right)\right), & n = 2, \\ \exp\left(\left(\log^{(n-2)}(t)\right)^{-\alpha} \exp\left(t(\log^{(n-2)}(t))^{-\alpha}\right)\right), & n \geq 3, \end{cases}$$

for type II. In the following sections, we make the above heuristics rigorous.

4 Preparations

In this section, we collect some tools to be used in the proofs of the Theorems 1 and 2. To be self-contained, we begin by restating Lemma 2 of Ref. [1] without repeating the proof.

Lemma 4.1. *On survival, $(W_t)_{t\geq 1}$ is almost surely an unbounded sequence.*

Other than in Ref. [1] the gap between the generation where a mutant type first appears and the generation where it may become dominant is unbounded. Therefore we need tight bounds on the Galton-Watson process with Poisson offspring distribution, which become the focus of the rest of this section. We prepare this with some bounds on the Poisson series.

Lemma 4.2. *If $0 < b < 1$, $\theta > 1$, $\lfloor b\theta \rfloor \geq 1$, and $(1-b)\theta \geq 1$, then*

$$\sum_{m=0}^{\lfloor b\theta \rfloor} e^{-\theta} \frac{\theta^m}{m!} \leq \theta e^{-\theta(1-b+b\log b)}.$$

Proof. Let $\ell := \lfloor b\theta \rfloor$ and $a_m := \theta^m/m!$. Note that $\ell \le b\theta \le \theta - 1$ by the assumption. Since $a_m/a_{m-1} = \theta/m$, we have $a_m \le a_\ell$ for all $m \le \ell < \theta$ and, therefore,

$$\sum_{m=0}^{\ell} \frac{\theta^m}{m!} \le (\ell+1)\frac{\theta^\ell}{\ell!} \le \theta \frac{\theta^\ell}{\ell!}.$$

Using $m! \ge m^m e^{-m}$ $(m \ge 1)$, we find $\log \frac{\theta^\ell}{\ell!} \le \ell \log\theta - \ell\log\ell + \ell$. Observing that $x\log\theta - x\log x + x$ is an increasing function in the region $0 < x < \theta$, we finally have

$$\sum_{m=0}^{\ell} e^{-\theta}\frac{\theta^m}{m!} \le \theta e^{-\theta + b\theta\log\theta - b\theta\log(b\theta) + b\theta} = \theta e^{-\theta(1-b+b\log b)},$$

as claimed. □

Lemma 4.3. *If $B > 1$ and $\theta > 0$, then*

$$\sum_{k=\lceil B\theta \rceil}^{\infty} e^{-\theta}\frac{\theta^k}{k!} \le \frac{B}{B-1} e^{-\theta(1-B+B\log B)}.$$

Proof. Let $m := \lceil B\theta \rceil$. Since $(m+k)! \ge m! m^k$ and $\theta/m \le 1/B < 1$, we have

$$\sum_{k=m}^{\infty}\frac{\theta^k}{k!} \le \frac{\theta^m}{m!}\sum_{k=0}^{\infty}\left(\frac{\theta}{m}\right)^k = \frac{\theta^m}{m!}\frac{1}{1-(\theta/m)} \le \frac{B}{B-1}e^{m\log\theta - m\log m + m} \le \frac{B}{B-1}e^{B\theta - B\theta\log B},$$

where we have used $m! \ge m^m e^{-m}$ and that $x\log\theta - x\log x + x$ is a decreasing function in the region $\theta < x$. Multiplying by $e^{-\theta}$, we get the desired inequality. □

Definition. By $(\mathcal{X}_t)_{t\ge 0}$, we mean a classical Galton-Watson process with Poisson offspring number distribution with mean θ, starting in generation 0 with a single individual.

Remark 4.1. *Conditioned on $\mathcal{X}_{t-1} = m$ for a nonnegative integer m, $\mathcal{X}_t$ is a Poisson-distributed random variable with mean $m\theta$.*

Lemma 4.4. *If $0 < b < 1$, $\theta \ge f \ge 1/(1-b+b\log b)$, and $x \ge 1$, then,*

$$\mathbb{P}(\mathcal{X}_t \ge bxf | \mathcal{X}_{t-1} \ge x) \ge 1 - xfe^{-xf(1-b+b\log b)}. \tag{18}$$

Proof. By assumption, $(1-b)f \ge 1$. If $m \ge x$, Remark 4.1 with Lemma 4.2 gives

$$\mathbb{P}(\mathcal{X}_t < bxf|\mathcal{X}_{t-1} = m) \le \mathbb{P}(\mathcal{X}_t < bm\theta|\mathcal{X}_{t-1} = m) \le \sum_{k=0}^{\lfloor bm\theta \rfloor} e^{-m\theta}\frac{(m\theta)^k}{k!} \le m\theta e^{-m\theta(1-b+b\log b)}.$$

Since $ze^{-zc} \le ye^{-yc}$ for all $z \ge y \ge 1/c > 0$ and $m\theta \ge xf \ge 1/(1-b+b\log b) > 0$, we get

$$\mathbb{P}(\mathcal{X}_t < bxf|\mathcal{X}_{t-1} = m) \le xfe^{-xf(1-b+b\log b)},$$

which does not depend on m as long as $m \ge x$. Now, the proof is completed. □

Lemma 4.5. *Let $A_t := \{a_t \le \mathcal{X}_t \le b_t\}$, where $0 \le a_t \le b_t - 1 \le \infty$ for all $t \ge 0$. Let $E_t := \bigcap_{k=\tau}^{t} A_k$ for $0 \le \tau < t$. Assume $\mathbb{P}(A_\tau) > 0$ and $\mathbb{P}(A_t|\mathcal{X}_{t-1} = m) \ge f_t > 0$, where m is any integer satisfying $a_{t-1} \le m \le b_{t-1}$ and f_t depends on a_t, b_t, a_{t-1}, and b_{t-1} but not on m. Then*

$$\mathbb{P}(E_t) \ge \mathbb{P}(A_\tau)\prod_{k=\tau+1}^{t} f_k.$$

Proof. For $t = \tau + 1$, the proof is trivial. So we assume $t \geq \tau + 2$. Note that

$$E_t = A_t \cap A_{t-1} \cap E_{t-2} = \bigcup_{m=\lceil a_{t-1} \rceil}^{\lfloor b_{t-1} \rfloor} \left(A_t \cap \{\mathcal{X}_{t-1} = m\} \cap E_{t-2} \right).$$

Using the countable additivity of the probability measure and the Markov property of $\mathcal{X}_t$,

$$\begin{aligned}
\mathbb{P}(E_t) &= \sum_{m=\lceil a_{t-1} \rceil}^{\lfloor b_{t-1} \rfloor} \mathbb{P}(A_t | \mathcal{X}_{t-1} = m) \mathbb{P}\left(\{\mathcal{X}_{t-1} = m\} \cap E_{t-2}\right) \\
&\geq f_t \sum_{m=\lceil a_{t-1} \rceil}^{\lfloor b_{t-1} \rfloor} \mathbb{P}\left(\{\mathcal{X}_{t-1} = m\} \cap E_{t-2}\right) = f_t \mathbb{P}(E_{t-1}).
\end{aligned}$$

Iterating the above inequality, we get the desired inequality. □

Lemma 4.6. *If $0 < b < 1$, $\theta \geq f \geq 1/(1 - b + b \log b)$, and $bf > 1$, then*

$$\mathbb{P}\left(\mathcal{X}_t \geq b^t f^t \text{ for all } t \geq \tau | \mathcal{X}_\tau \geq b^\tau f^\tau\right) \geq 1 - f\left(1 + \tfrac{1}{\log(bf)}\right) e^{-f(1-b+b\log b)},$$

for any nonnegative integer τ. Note that the right hand side does not depend on τ.

Proof. For any event E, we write $\mathbb{P}_c(E) := \mathbb{P}(E | \mathcal{X}_\tau \geq b^\tau f^\tau)$ in this proof. Define

$$A_t := \{\mathcal{X}_t \geq b^t f^t\}, \quad C_t := \bigcap_{k=\tau+1}^{t} A_k, \quad C := \bigcap_{k=\tau+1}^{\infty} A_k.$$

Note that

$$\mathbb{P}_c(A_\tau) = 1, \quad \mathbb{P}_c\left(\mathcal{X}_t \geq b^t f^t \text{ for all } t \geq \tau\right) = \mathbb{P}_c(C) = \lim_{t \to \infty} \mathbb{P}_c(C_t).$$

Using (18) with $x \mapsto (bf)^{t-1}$, we have

$$\mathbb{P}(A_t | A_{t-1}) \geq 1 - f(bf)^{t-1} \exp\left[-f(bf)^{t-1}(1 - b + b\log b)\right] =: 1 - d_t.$$

By Lemma 4.5 we can write

$$\mathbb{P}_c(C) = \lim_{t\to\infty} \mathbb{P}_c(C_t) \geq \prod_{t=\tau+1}^{\infty} (1 - d_t) \geq 1 - \sum_{t=\tau+1}^{\infty} d_t \geq 1 - \sum_{t=1}^{\infty} d_t.$$

Since $(c^{t-1} \exp(-ac^{t-1}))_{t \geq 1}$ is a decreasing sequence for $a \geq 1$ and $c > 1$, we have

$$\begin{aligned}
&\sum_{t=1}^{\infty} c^{t-1} \exp\left(-ac^{t-1}\right) = e^{-a} + \sum_{t=2}^{\infty} c^{t-1} \exp\left(-ac^{t-1}\right) \\
&\leq e^{-a} + \int_1^\infty c^{t-1} \exp\left(-ac^{t-1}\right) dt = \left(1 + \frac{1}{a \log c}\right) e^{-a} \leq \left(1 + \frac{1}{\log c}\right) e^{-a}.
\end{aligned} \tag{19}$$

Plugging $c = bf$ and $a = f(1 - b + b \log b)$ into (19), we have the desired result. □

Remark 4.2. *If we restrict the condition of parameters in Lemma 4.6 to be $bf \geq e$ and $0 < b \leq b_c < 1/2$, where b_c satisfies $1 - b_c + b_c \log b_c = 1/2$, then we can use*

$$\mathbb{P}\left(\mathcal{X}_t \geq b^t f^t \text{ for all } t \geq \tau | \mathcal{X}_\tau \geq b^\tau f^\tau\right) \geq 1 - 2fe^{-f/2}. \tag{20}$$

Lemma 4.7. *If $B > 1$, $f \ge \theta > 0$, and $x \ge 1$, then*

$$\mathbb{P}(\mathcal{X}_t \le Bxf|\mathcal{X}_{t-1} \le x) \ge 1 - \frac{B}{B-1}e^{-xfB(\log B-1)}.$$

Proof. Set $\mathcal{X}_{t-1} = m$. If $m = 0$, the above inequality is trivially true. So we only consider $1 \le m \le x$. Let $B' := Bxf/(m\theta) \ge B$. Then, Remark 4.1 together with Lemma 4.3 gives

$$\mathbb{P}(\mathcal{X}_t > Bxf|\mathcal{X}_{t-1} = m) = \mathbb{P}(\mathcal{X}_t > B'm\theta|\mathcal{X}_{t-1} = m) \le \sum_{k=\lceil B'm\theta \rceil}^{\infty} \frac{(m\theta)^k}{k!} \le \frac{B'}{B'-1}e^{-m\theta B'(\log B'-1)},$$

where we have used $e^{-y} \le 1$ for $y \ge 0$. Since $xfB = m\theta B'$, $\log B' \ge \log B$, and $y/(y-1)$ is a decreasing function of $y > 1$, we have

$$\mathbb{P}(\mathcal{X}_t > Bxf|\mathcal{X}_{t-1} = m) \le \frac{B}{B-1}e^{-xfB(\log B-1)},$$

which is valid for any $m \le x$. Now the proof is completed. □

Remark 4.3. *In case $B \ge e^2 > 2$, we can use*

$$\mathbb{P}(\mathcal{X}_t \le Bxf|\mathcal{X}_{t-1} \le x) \ge 1 - 2e^{-xf}. \tag{21}$$

We next describe the distribution of W_t, conditioned on $\Xi(t) = N$.

Lemma 4.8. *For any $x \ge 0$,*

$$\mathbb{P}(W_t \le x|\Xi(t) = N) = (1 - \beta G(x))^N.$$

Proof. First fix a positive integer m and by $W_t^{(m)}$ is denoted the largest of m independently sampled fitnesses with convention $W_t^{(0)} = 0$. Then $\mathbb{P}(W_t^{(m)} \le x) = (1 - G(x))^m$. Let q_m be the probability that m mutants arise out of N. Then,

$$\begin{aligned}\mathbb{P}(W_t \le x|\Xi(t) = N) &= \sum_{m=0}^{N} \mathbb{P}(W_t^{(m)} \le x)q_m = \sum_{m=0}^{N}(1 - G(x))^m q_m \\ &= \sum_{m=0}^{N}(1 - G(x))^m \binom{N}{m}\beta^m(1-\beta)^{N-m} = (1 - \beta G(x))^N,\end{aligned}$$

as claimed. □

Remark 4.4. *In case $X(t) \le \Xi(t) \le y$, we will use the inequality*

$$\mathbb{P}(W_t \le x|\Xi(t) \le y) \ge 1 - \beta y G(x), \tag{22}$$

where we have used $(1-z)^m \ge 1 - mz$ for $0 \le z \le 1$ and $m \ge 1$. In case $\Xi(t) \ge X(t) \ge y \ge 0$, we will use the inequality

$$\mathbb{P}(W_t \ge x|X(t) \ge y) \ge 1 - e^{-\beta y G(x)}, \tag{23}$$

where we have used $e^{-yz} \ge (1-z)^y$ for $0 \le z \le 1$.

5 Proof of Theorem 1

We first provide a heuristic argument for a more accurate estimate of W_t than in Section 3. As in Theorem 1 we assume

$$\log X(t) \approx \frac{t}{\alpha} \log^{(n)}(t) .$$

Then, we approximate

$$\log^{(n)}(X(t)) \approx \log^{(n-1)}(t) \left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}} .$$

Now using the mean-field type approximation $g_I(W_t) \approx \log^{(n)}(1/G(W_t)) \approx \log^{(n)}(X(t))$, we get an approximate W_t by a solution x of the equation

$$x^\alpha L(x) = g_I(x) = \log^{(n-1)}(t) \left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}} .$$

We can find an approximate solution of the above equation as

$$\begin{aligned} x &= L(x)^{-1/\alpha} \left(\log^{(n-1)}(t)\right)^{1/\alpha} \left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}/\alpha} \\ &\approx \left(\log^{(n-1)}(t)\right)^{1/\alpha} \left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}/\alpha} \left[L\left((\log^{(n-1)}(t))^{1/\alpha}\right)\right]^{-1/\alpha} = u_n(t). \end{aligned}$$

where we have used **(A2)**. By construction, we have

$$g_I(u_n(t)) \sim \log^{(n-1)}(t) \left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}} , \tag{24}$$

which will play an important role in proving Theorem 1. In the proof, no distinction between MMM and FMM is necessary. For the proof, we begin with estimating $G(W_t)$ using an inequality relating the iterated exponential function $\exp^{(n)}(x)$ and the iterated logarithm $\log^{(n)}(x)$.

Lemma 5.1. *For any positive integer n and for any positive x,*

$$\exp^{(n)}\left(x \log^{(n)}(t)\right) \geq t^x, \tag{25}$$

as long as $\log^{(n)}(t) > 1$.

Proof. For $n = 1$, the inequality is trivially valid as an equality. Now assume that the inequality is satisfied for $n = \ell$. Consider t with $\log^{(\ell+1)}(t) > 1$, which gives $\log^{(\ell)}(t) > e > 1$ and $t > 1$. Abbreviate $y := (\log^{(\ell)}(t))^{x-1}$ for $x > 0$. Since $\log^{(\ell)}(t) > e$, we have $y > e^{x-1} > x$. By assumption, we have

$$\exp^{(\ell+1)}\left(x \log^{(\ell+1)}(t)\right) = \exp^{(\ell)}\left(\left(\log^{(\ell)}(t)\right)^x\right) = \exp^{(\ell)}\left(y \log^{(\ell)}(t)\right) \geq t^y \geq t^x,$$

so that the claimed inequality is also valid for $n = \ell + 1$. Induction completes the proof. □

Lemma 5.2. *Fix ε such that $0 < \varepsilon < \frac{2\alpha}{3\alpha+8+\sqrt{\alpha^2+48\alpha+64}}$, and let $\varepsilon_1 := \frac{4\varepsilon}{\alpha(1-2\varepsilon)(1-\varepsilon)-4\varepsilon}$. Then there is τ_0 (depending on n) such that, for all $t \geq \tau_0$,*

$$\begin{aligned} \log G((1-\varepsilon_1)u_n(t)) &\geq -\frac{1-2\varepsilon}{\alpha} t \log^{(n)}(t) , \\ \log G((1+\varepsilon_1)u_n(t)) &\leq -\frac{1+2\varepsilon}{\alpha} t \log^{(n)}(t) . \end{aligned}$$

Proof. First note that $\varepsilon < 1/2$ for any $\alpha > 0$, $0 < \varepsilon_1 < 1$, and

$$(1-2\varepsilon)\left[1+\alpha(1-\varepsilon)\tfrac{\varepsilon_1}{1+\varepsilon_1}\right] = 1+2\varepsilon, \quad \tfrac{1+2\varepsilon}{1+\alpha\varepsilon_1(1-\varepsilon)} \le 1-2\varepsilon. \tag{26}$$

By (24), there is τ_1 such that

$$\tfrac{1-2\varepsilon}{1-\varepsilon}\log^{(n-1)}(t)\left(\tfrac{\log t}{\alpha}\right)^{\delta_{n,1}} \le g_I(u_n(t)) \le \tfrac{1+2\varepsilon}{1+\varepsilon}\log^{(n-1)}(t)\left(\tfrac{\log t}{\alpha}\right)^{\delta_{n,1}}, \tag{27}$$

for all $t \ge \tau_1$. Now we show that there is τ_2 such that, for all $t \ge \tau_2$,

$$\begin{aligned} \log^{(n-1)}\left(\tfrac{1+2\varepsilon}{\alpha}\, t\log^{(n)}(t)\right) &\le (1+2\varepsilon)\log^{(n-1)}(t)\left(\tfrac{\log t}{\alpha}\right)^{\delta_{n,1}}, \\ \log^{(n-1)}\left(\tfrac{1-2\varepsilon}{\alpha}\, t\log^{(n)}(t)\right) &\ge (1-2\varepsilon)\log^{(n-1)}(t)\left(\tfrac{\log t}{\alpha}\right)^{\delta_{n,1}}. \end{aligned} \tag{28}$$

For $n = 1$, this is obvious. For $n \ge 2$, existence of τ_2 follows from

$$\lim_{t\to\infty} \tfrac{1}{\log^{(n-1)}(t)} \log^{(n-1)}\left(\tfrac{1\pm 2\varepsilon}{\alpha}\, t\log^{(n)}(t)\right) = 1.$$

By the mean value theorem, there is $\varepsilon_\pm$ such that $0 \le \varepsilon_\pm \le \varepsilon_1$ and

$$g_I((1\pm\varepsilon_1)u_n(t)) = g_I(u_n(t)) \pm \varepsilon_1 u_n(t)\, \frac{dg_I}{dx}\bigg|_{x=(1\pm\varepsilon_\pm)u_n(t)}. \tag{29}$$

We do not make the t-dependence of $\varepsilon_\pm$ explicit, as in the following we will only use the inequality $0 < \varepsilon_\pm < \varepsilon_1$. By **(A4)** with $j = \gamma = 1$ and (7),

$$\alpha = \lim_{x\to\infty} \frac{x}{g_I(x)} \frac{dg_I(x)}{dx} = \lim_{x\to\infty} \frac{d\log g_I(x)}{d\log x}.$$

Hence there is x_1 such that, for all $x' \ge x \ge x_1$, we have $g_I(x') \ge g_I(x)$ and

$$\alpha(1-\varepsilon)\frac{g_I(x)}{x} \le \frac{dg_I}{dx}, \tag{30}$$

and

$$\exp\left(-\exp^{(n-1)}((1+\varepsilon)g_I(x))\right) \le G(x) \le \exp\left(-\exp^{(n-1)}((1-\varepsilon)g_I(x))\right). \tag{31}$$

Hence, if $(1-\varepsilon_1)u_n(t) > x_1$, then we have

$$\begin{aligned} g_I((1+\varepsilon_1)u_n(t)) &\ge g_I(u_n(t)) + \varepsilon_1 u_n(t)\alpha(1-\varepsilon)\frac{g_I((1+\varepsilon_+)u_n(t))}{(1+\varepsilon_+)u_n(t)} \\ &\ge g_I(u_n(t)) + \frac{\varepsilon_1}{1+\varepsilon_1}\alpha(1-\varepsilon)g_I((1+\varepsilon_+)u_n(t)) \\ &\ge g_I(u_n(t))\left[1+\alpha(1-\varepsilon)\frac{\varepsilon_1}{1+\varepsilon_1}\right], \end{aligned}$$

where we have used (29), (30), $1/(1+\varepsilon_+) \ge 1/(1+\varepsilon_1)$ and $g_I((1+\varepsilon_+)u_n(t)) \ge g_I(u_n(t))$; and

$$\begin{aligned} g_I((1-\varepsilon_1)u_n(t)) &\le g_I(u_n(t)) - \varepsilon_1 u_n(t)\alpha(1-\varepsilon)\frac{g_I((1-\varepsilon_-)u_n(t))}{(1-\varepsilon_-)u_n(t)} \\ &\le g_I(u_n(t)) - \varepsilon_1\alpha(1-\varepsilon)g_I((1-\varepsilon_1)u_n(t)), \end{aligned} \tag{32}$$

where we have used (29), (30), $-1/(1-\varepsilon_-) \le -1$ and $-g_I((1-\varepsilon_-)u_n(t)) \le -g_I((1-\varepsilon_1)u_n(t))$. We can rewrite (32) as $g_I((1-\varepsilon_1)u_n(t)) \le [1+\alpha\varepsilon_1(1-\varepsilon)]^{-1} g_I(u_n(t))$. Therefore, there is τ_3 such that for all $t \ge \tau_3$ we have

$$\begin{aligned}(1-\varepsilon)g_I((1+\varepsilon_1)u_n(t)) &\ge (1-\varepsilon)g_I(u_n(t))\left[1+\alpha(1-\varepsilon)\frac{\varepsilon_1}{1+\varepsilon_1}\right]\\ &\ge (1-2\varepsilon)\left[1+\alpha(1-\varepsilon)\frac{\varepsilon_1}{1+\varepsilon_1}\right]\log^{(n-1)}(t)\left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}}\\ &= (1+2\varepsilon)\log^{(n-1)}(t)\left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}} \ge \log^{(n-1)}\left(\tfrac{1+2\varepsilon}{\alpha}t\log^{(n)}(t)\right),\end{aligned} \tag{33}$$

and

$$\begin{aligned}(1+\varepsilon)g_I((1-\varepsilon_1)u_n(t)) &\le (1+\varepsilon)g_I(u_n(t))\left[1+\alpha\varepsilon_1(1-\varepsilon)\right]^{-1}\\ &\le \frac{1+2\varepsilon}{1+\alpha\varepsilon_1(1-\varepsilon)}\log^{(n-1)}(t)\left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}}\\ &\le (1-2\varepsilon)\log^{(n-1)}(t)\left(\frac{\log t}{\alpha}\right)^{\delta_{n,1}} \le \log^{(n-1)}\left(\tfrac{1-2\varepsilon}{\alpha}t\log^{(n)}(t)\right),\end{aligned} \tag{34}$$

where we have used (26), (27), and (28). To sum up, there is τ_0 such that for all $t \ge \tau_0$,

$$\log G((1+\varepsilon_1)u_n(t)) \le -\tfrac{1+2\varepsilon}{\alpha}t\log^{(n)}(t)\,,$$

where we have used (31) and (33); and

$$\log G((1-\varepsilon_1)u_n(t)) \ge -\tfrac{1-2\varepsilon}{\alpha}t\log^{(n)}(t)\,,$$

where we have used (31) and (34). Now, the proof is completed. □

Lemma 5.3. *Assume $X(0) < \infty$ and $Q_0 < \infty$. Fix ε and ε_1 as in Lemma 5.2 and let, for $t \ge 0$,*

$$A_t := \left\{\log \Xi(t) \le \tfrac{1+\varepsilon}{\alpha}(t+m)\log^{(n)}(t+m)\right\}, \quad E_t := \{W_t \le (1+\varepsilon_1)\,u_n(t+m)\}\,,$$

where m is assumed large enough for the definition to make sense. We use the convention $\log 0 = -\infty$ throughout the paper. We define a sequence of events $(D_t)_{t\ge 0}$ iteratively as

$$D_0 = A_0 \cap E_0, \quad D_t = A_t \cap E_t \cap D_{t-1}.$$

Let $D := \bigcap_{t=0}^{\infty} D_t$. Then,

$$\lim_{m\to\infty} \mathbb{P}(D) = 1.$$

Proof. Since $\liminf_{t\to\infty} u_n(t) = \infty$ and $\log^{(n)}(t)$ is an unbounded and increasing function, there is t_1 such that $u_n(m) \ge 1$, $(1+\varepsilon)m\log^{(n)}(m) \ge \alpha(m+1)$, $(1+\varepsilon)m\log^{(n)}(m) \ge \alpha\log X(0)$ and $(1+\varepsilon_1)u_n(m) \ge Q_0$ for all $m > t_1$. Let

$$\begin{aligned}H(x) &:= \frac{1+\varepsilon}{\alpha}(x+1)\log^{(n)}(x+1) - \frac{1+\varepsilon}{\alpha}x\log^{(n)}(x) - \log(u_n(t)) - \log(1+\varepsilon_1)\\ &= \frac{x+1}{\alpha'}\left[\log^{(n)}(x+1) - \log^{(n)}(x)\right] + \varepsilon\frac{\log^{(n)}(x)}{\alpha} - \log\left(\omega_W(\log^{(n)}(x))\right) - \log(1+\varepsilon_1),\end{aligned}$$

where $\alpha' := \alpha/(1+\varepsilon)$. Since $\liminf_{x\to\infty} H(x) = \infty$, there is t_2 such that $H(x) > 2$ for all $x > t_2$. By Lemma 5.2, we can choose t_3 such that

$$\log G((1+\varepsilon_1)u_n(x)) \le -\frac{1+2\varepsilon}{\alpha} x \log^{(n)}(x), \tag{35}$$

for all $x > t_3$. From now on, we only consider large m such that $m > t_0 := \max\{t_1, t_2, t_3\}$. For convenience, we define $\tau_t := t + m$. Let $E'_t := \{Q_t \le (1+\varepsilon_1)u_n(\tau_t)\}$. Since $E_0 = E'_0$ and $E_{t+1} \cap E'_t = E'_{t+1} \cap E'_t$ even though E'_t can be a proper subset of E_t, we have

$$\bigcap_{k=0}^{t} E_k = \bigcap_{k=0}^{t} E'_k. \tag{36}$$

We have, for $k \ge 1$, that

$$\mathbb{P}\left(A_k | A_{k-1} \cap E'_{k-1}\right) \ge 1 - 2\exp\left(-e^{\tau_k}\right) =: 1 - \xi_k,$$

where we have used (21) with $f \mapsto (1+\varepsilon_1)u_n(\tau_{k-1}) \ge 1$, $x \mapsto \exp(\frac{1+\varepsilon}{\alpha}\tau_{k-1}\log^{(n)}(\tau_{k-1})) \ge e^{\tau_k}$, and $B \mapsto e^{H(\tau_{k-1})} \ge e^2$ and the fact $S_t \le Q_t$.

Observe that $\mathbb{P}(D_t) = \mathbb{P}(E_t | A_t \cap D_{t-1})\mathbb{P}(A_t|D_{t-1})\mathbb{P}(D_{t-1})$. Using Lemma 4.5 and (36), we have

$$\mathbb{P}\left(A_k | D_{k-1}\right) \ge 1 - \xi_k$$

Since W_k is purely determined by $\Xi(k)$, E_k is independent of D_{k-1} and, accordingly, we have

$$\mathbb{P}(E_k | A_k \cap D_{k-1}) = \mathbb{P}(E_k|A_k) \ge 1 - \beta \exp\left(-\frac{\varepsilon}{\alpha}\tau_k \log^{(n)}(\tau_k)\right) =: 1 - \eta_k,$$

where we have used (22) with $\alpha \log y \mapsto (1+\varepsilon)\tau_k \log^{(n)}(\tau_k)$, $x \mapsto (1+\varepsilon_1)u_n(\tau_k)$, and (35). Therefore,

$$\mathbb{P}(D) \ge \prod_{k=1}^{\infty}(1-\xi_k)(1-\eta_k) \ge 1 - \sum_{k=1}^{\infty}(\xi_k + \eta_k). \tag{37}$$

Note that $\lim_{m\to\infty}(\xi_k + \eta_k) = 0$. Since $\lim_{k\to\infty}(\xi_k + \eta_k)\tau_k^2 = 0$, there is a constant c that is independent of m such that $\xi_k + \eta_k \le c\tau_k^{-2} \le ck^{-2}$ for all k. Hence, the series in (37) converges uniformly for all $m > t_0$ and, therefore, $\lim_{m\to\infty} \mathbb{P}(D) = 1$, which completes the proof. □

Lemma 5.4 (Upper bound). *If $X(0) < \infty$ and $Q_0 < \infty$, then almost surely,*

$$\limsup_{t\to\infty} \frac{\log X(t)}{t \log^{(n)}(t)} \le \frac{1}{\alpha}, \quad \limsup_{t\to\infty} \frac{W_t}{u_n(t)} \le 1.$$

Proof. Choose ε and ε_1 as in Lemma 5.2. Let

$$C(\varepsilon) := \left\{\limsup_{t\to\infty} \frac{\log X(t)}{t \log^{(n)}(t)} \le \frac{1+\varepsilon}{\alpha}\right\},$$

$$\tilde{C}(m,\varepsilon) := \left\{\log X(t) \le (1+\varepsilon)\frac{(t+m)\log^{(n)}(t+m)}{\alpha} \text{ for all } t\right\}.$$

We use D in Lemma 5.3 with m to be the same meaning as in this lemma. Since $D \subset \tilde{C}(m,\varepsilon) \subset C(\varepsilon)$ for any $m > 0$, Lemma 5.3 gives $\mathbb{P}(C(\varepsilon)) = 1$. Defining $E = \bigcap_{t=1}^{\infty} E_t$ we get $\lim_{m\to\infty} \mathbb{P}(E) = 1$, because $D \subset E$. Therefore,

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{W_t}{u_n(t)} \le 1 + \varepsilon_1\right) \ge \lim_{m\to\infty} \mathbb{P}(E) = 1.$$

Since ε is arbitrary, the proof is completed. □

Definition (Initial condition for Lemma 5.6). Choose ε as in Lemma 5.2. Let

$$\alpha_1 := \alpha\left(1-\frac{\varepsilon}{2}\right)^{-1}, \quad f_k := (1-\beta)u_n(k), \quad b_k := \frac{1}{1-\beta}\exp\left(-\frac{\varepsilon}{2\alpha}\log^{(n)}(k)\right),$$
$$b_k f_k = \exp\left(\frac{\log^{(n)}(k)}{\alpha_1} + \log\left(\omega_W\left(\log^{(n)}(k)\right)\right)\right), \tag{38}$$

where k is assumed sufficiently large in order for the definition to make sense. We also define

$$H(m,x) := \frac{\log^{(n)}(m)}{\alpha_1}(x-m) + (x-m)\log\left(\omega_W(\log^{(n)}(m))\right),$$
$$h(m,x) := H(m,x) - \frac{1-\varepsilon}{\alpha}x\log^{(n)}(x),$$
$$\tau_j(m) := \exp^{(n)}\left((1+j\varepsilon_2)\log^{(n)}(m)\right), \quad \varepsilon_2 := \frac{\varepsilon}{8(1-\varepsilon)} < \frac{\varepsilon}{4}, \tag{39}$$

where $j = 1, 2$ and we assume $\log^{(n)}(m) > 1$, $\omega_W(\log^{(n)}(m)) > 0$, and $x > m$. Note that $(1-\varepsilon/2)/(1+2\varepsilon_2) > 1-\varepsilon$. Since

$$(b_m f_m)^{x-m}\exp\left(-\frac{1-\varepsilon}{\alpha}x\log^{(n)}(x)\right) = e^{h(m,x)},$$

$h(m,x) \geq 0$ implies $(b_m f_m)^{x-m} \geq \exp\left(\frac{1-\varepsilon}{\alpha}x\log^{(n)}(x)\right)$.

We choose an integer k_0 as in Lemma 5.5. Once k_0 is fixed, we define an initial condition for any integer $t_0 \geq k_0$. In generation 0, there are $t_0 - k_0 + 1$ different mutant types with fitness $F_k = f_k/(1-\beta)$ ($k_0 \leq k \leq t_0$) and the number $M_k(0)$ of individuals with fitness F_k is $M_k(0) = \lceil f_k^{t_0-k}\rceil \geq (b_k f_k)^{t_0-k}$. We denote the number of nonmutated descendants of $M_k(0)$ in generation t by $M_k(t)$.

For convenience, we denote the largest fitness among mutants at generation $k \geq 1$ by F_{k+t_0} and its nonmutated descendants at generation $t \geq k$ by $M_{k+t_0}(t)$. Note that $W_k = F_{k+t_0}$ and $N_k(t) = M_{k+k_0}(t)$ for $k \geq 1$. We set $F_{k+t_0} = 0$ if there are no new mutants at generation k. If $F_{k+t_0} = 0$, we write $M_{k+t_0}(t) = 0$ for all t. If $F_{k+t_0} > 0$, we set $M_{k+t_0}(k) = 1$. That is, even if there are many mutants with the same largest fitness F_{k+t_0}, which may frequently happen in the MMM if discrete fitness values are allowed to be sampled, $M_{k+t_0}(t)$ only concerns descendants of a single individual among them. Finally, we define

$$\mathcal{Y}(t) := \sum_{k=k_0}^{t+t_0} M_k(t).$$

Note that $\mathcal{Y}(t) \leq X(t)$ and equality holds for the FMM.

Lemma 5.5. *For b_k, f_k in* (38) *and for H, h, τ_1, τ_2 in* (39)*, there is an integer k_0, which is larger than $\exp^{(n)}(1)$, such that for all $m \geq k_0$*

(Condition 1) $0 < b_m < b_c$, $(1 - b_m + b_m \log b_m)f_m > 1$, *and* $b_m f_m > e$ *(see Remark 4.2 for the motivation of this condition);*

(Condition 2) $h(m,x) > 0$ *with any x satisfying* $\tau_1(m) \leq x \leq \tau_2(m)$.

Proof. It is obvious that there is an integer k_1 such that (**Condition 1**) is satisfied for all $m \geq k_1$. For any sequence $(x_m)_{m \geq \lceil \exp^{(n)}(1) \rceil}$ such that $\tau_1(m) \leq x_m \leq \tau_2(m)$, we have

$$\liminf_{m \to \infty} \frac{\alpha H(m, x_m)}{x_m \log^{(n)}(x_m)} \geq \frac{1-\varepsilon/2}{1+2\varepsilon_2} \liminf_{m \to \infty} \frac{(x_m - m)\log^{(n)}(m)}{x_m \log^{(n)}(m)} = \frac{1-\varepsilon/2}{1+2\varepsilon_2} > 1 - \varepsilon, \tag{40}$$

where we have used $\log^{(n)}(x_m) \leq (1+2\varepsilon_2)\log^{(n)}(m)$ by assumption, $x_m \geq \tau_1(m) \geq m^{1+\varepsilon_2}$ for $m \geq \exp^{(n)}(1)$ (Lemma 5.1), and ω_W is a slowly varying function. Therefore, there is an integer k_2 such that $h(m,x) > 0$ for all $m \geq k_2$ with any x satisfying $\tau_1(m) \leq x \leq \tau_2(m)$. Now we set $k_0 = \max\{\lceil \exp^{(n)}(1) \rceil, k_1, k_2\}$ and the proof is completed. □

Remark 5.1. *Inequality* (40) *is valid even if we relax the lower bound of x_m as long as $\lim_{m\to\infty} m/x_m = 0$. For example, replacing $\tau_1(m)$ by $m\sqrt{\log m}$ still gives $h(m,x) > 0$ for sufficiently large m.*

Lemma 5.6. *We fix ε and ε_1 as in Lemma 5.2. We also use the initial conditions defined above with $t_0 \geq k_0$ and define*

$$E_t := \left\{ \log \mathcal{Y}(t) \geq \frac{1-\varepsilon}{\alpha}(t+t_0)\log^{(n)}(t+t_0) \right\}, \quad E := \bigcap_{t=1}^{\infty} E_t,$$

$$J_t := \{W_t \geq (1-\varepsilon_1)u_n(t+t_0)\}, \quad J := \bigcap_{t=1}^{\infty} J_t.$$

Then,

$$\lim_{t_0 \to \infty} \mathbb{P}(E) = \lim_{t_0 \to \infty} \mathbb{P}(J) = 1.$$

Proof. We define a sequence $(m_\ell)_{\ell \geq 0}$ as

$$m_\ell := \left\lfloor \exp^{(n-1)}\left(\ell + \log^{(n-1)}\left(\tfrac{t_0}{\log t_0}\right)\right)\right\rfloor.$$

We first work out how large t_0 should be. Obviously, there exists t_1 such that $k_0 < m_0 < t_0$ for all $t_0 > t_1$. Since

$$\lim_{t_0 \to \infty} \frac{t_0}{\tau_2(m_0)} = 0, \quad \lim_{t_0 \to \infty} \frac{\alpha H(m_0, t_0+1)}{(t_0+1)\log^{(n)}(t_0+1)} = 1 > 1 - \varepsilon,$$

there is t_2 such that $\tau_2(m_0) \geq t_0 + 1$ and $h(m_0, t_0+1) > 0$ for all $t_0 > t_2$. In fact, $h(m_0, x) > 0$ for all $t_0 + 1 \leq x \leq \tau_2(m_0)$; see Remark 5.1. Since

$$\lim_{t_0 \to \infty} \frac{\log^{(n)}(m_\ell)}{\log^{(n)}(m_{\ell-1})} = \lim_{\ell \to \infty} \frac{\log^{(n)}(m_\ell)}{\log^{(n)}(m_{\ell-1})} = 1,$$

there is t_3 such that $\tau_2(m_{\ell-1}) \geq \tau_1(m_\ell)$ for all $t_0 > t_3$ and for all $\ell \geq 1$. By Lemma 5.2, there is t_4 such that

$$\log G\left((1-\varepsilon_1)u_n(t+t_0)\right) \geq -\frac{1-2\varepsilon}{\alpha}(t+t_0)\log^{(n)}(t+t_0), \tag{41}$$

for all $t_0 > t_4$ and for all $t \geq 0$. For later references, we define sequences $(\xi_\ell)_{\ell \geq 0}$ and $(\eta_t)_{t \geq 0}$ as

$$\eta_t := \exp\left(-\beta \exp\left(\frac{\varepsilon}{\alpha}(t+t_0)\log^{(n)}(t+t_0)\right)\right),$$
$$\xi_\ell := 2(1-\beta)u_n(m_\ell)\exp\left(-\tfrac{1-\beta}{2}u_n(m_\ell)\right).$$

Since $\log^{(n)}(x)$ is unbounded and increasing function for large x, there is t_5 such that $\varepsilon \log^{(n)}(t_0) \geq \alpha$ for all $t_0 > t_5$ and all $\ell \geq 0$. Therefore, we have $\eta_t \leq \exp(-\beta e^t)$, for all $t_0 > t_5$. This implies that $\sum_t \eta_t$ is uniformly convergent for all t_0 and therefore,

$$\lim_{t_0 \to \infty} \sum_{t=1}^{\infty} \eta_t = \sum_{t=1}^{\infty} \lim_{t_0 \to \infty} \eta_t = 0. \tag{42}$$

Since $4xe^{-x} \leq e^{-x/2}$ for $x \geq 10$, we have $\xi_\ell \leq \exp(-\frac{1-\beta}{4} u_n(m_\ell))$, for $(1-\beta)u_n(m_\ell) \geq 10$. Since

$$\lim_{x \to \infty} \frac{4\left(\log^{(n-1)}(x)\right)^{1/\alpha_1}}{(1-\beta)u_n(x)} = 0,$$

there is t_6 such that $(1-\beta)u_n(m_\ell) \geq 10$ for any $\ell \geq 0$ and

$$\frac{1-\beta}{4} u_n(m_\ell) \geq \left(\log^{(n-1)}(m_\ell)\right)^{1/\alpha_1} \geq \ell^{1/\alpha_1},$$

for all $t_0 > t_6$. Note that, under this assumption, we have $\xi_\ell \leq \exp\left(-\ell^{1/\alpha_1}\right)$, which shows that $\sum_{\ell=0}^{\infty} \xi_\ell$ converges uniformly for all large t_0 and therefore,

$$\lim_{t_0 \to \infty} \sum_{\ell=0}^{\infty} \xi_\ell = \sum_{\ell=0}^{\infty} \lim_{t_0 \to \infty} \xi_\ell = 0. \tag{43}$$

In the following, we assume $t_0 > \max\{t_1, t_2, t_3, t_4, t_5, t_6\}$.

Now we are ready for the proof. We first define two sequences $(a_\ell)_{\ell \geq 0}$ and $(u_\ell)_{\ell \geq 0}$ such that $a_0 = 0$, $a_\ell = \tau_1(m_\ell) - t_0$ for $\ell \geq 1$, and $u_\ell = \tau_2(m_\ell) - t_0$ for $\ell \geq 0$. Note that $a_{\ell+1} \leq u_\ell$ for all $\ell \geq 0$ and $m_\ell < t_0 + a_\ell$. Notice also that for $a_\ell \leq t < a_{\ell+1} \leq u_\ell$

$$(t + t_0 - m_\ell) \log(b_{m_\ell} f_{m_\ell}) \geq \frac{1-\varepsilon}{\alpha}(t+t_0) \log^{(n)}(t+t_0), \tag{44}$$

which also implies E_0 is an almost sure event. For $t \geq 0$, we define

$$A_t := \left\{ M_{m_{\ell'}} \geq \left(b_{m_{\ell'}} f_{m_{\ell'}}\right)^{t+t_0-m_{\ell'}} \right\},$$

where ℓ' is (uniquely) determined by the condition $a_{\ell'} \leq t < a_{\ell'+1}$. Note that $A_t \subset E_t$. Define

$$\tilde{J} := \bigcap_{m=m_0}^{t_0} \{F_m \geq f_m/(1-\beta)\}, \quad C_0 := A_0 \cap \tilde{J}, \quad C_t := J_t \cap A_t \cap C_{t-1}, \quad C := \bigcap_{t=1}^{\infty} C_t.$$

Note that $\tilde{J}$ and A_0 are sure events and so is C_0. Observe that

$$\mathbb{P}(C_t) = \mathbb{P}(J_t | A_t \cap C_{t-1}) \mathbb{P}(A_t | C_{t-1}) \mathbb{P}(C_{t-1}).$$

Since W_t is solely determined by $\Xi(t)$ and $\alpha \log \Xi(t) \geq (1-\varepsilon)(t+t_0)\log^{(n)}(t+t_0)$ in the event $A_t \cap C_{t-1}$, we have $\mathbb{P}(J_t | A_t \cap C_{t-1}) \geq 1 - \eta_t$, where we have used (23) with $\alpha \log y \mapsto (1-\varepsilon)(t+t_0)\log^{(n)}(t+t_0)$, $x \mapsto (1-\varepsilon_1)u_n(t)$ with (41). Therefore, we have

$$\mathbb{P}(C_t) \geq \left(\prod_{\tau=1}^{t} (1-\eta_\tau)\right) \prod_{\ell=0}^{\ell'} P_\ell, \quad P_\ell := \prod_{\tau=a_\ell}^{a_{\ell+1}-1} \mathbb{P}(A_\tau | C_{\tau-1}),$$

where we have used the fact that probability cannot be larger than 1.

Let us find the lower bound of P_ℓ. Assume $a_\ell \le \tau < a_{\ell+1}$. Note that A_τ is independent of J_k for $a_\ell \le k < \tau$ (this is because $m_\ell < a_\ell + t_0$) and of A_k for $k < a_\ell$ (this is because $M_m(t)$'s for different m's are mutually independent branching processes). Therefore,

$$\mathbb{P}(A_\tau | C_{\tau-1}) = \mathbb{P}\bigg(A_\tau \bigg| \Big(\bigcap_{k=a_\ell}^{\tau-1} A_k\Big) \cap J_{m_\ell - t_0}\bigg),$$

where $J_{m_\ell - t_0}$ for $m_\ell < t_0$ should be interpreted as $\tilde{J}$. By simple algebra, we get

$$\begin{aligned}
P_\ell &= \prod_{\tau=a_\ell}^{a_{\ell+1}-1} \mathbb{P}\bigg(A_\tau \bigg| \Big(\bigcap_{k=a_\ell}^{\tau-1} A_k\Big) \cap J_{m_\ell - t_0}\bigg) = \mathbb{P}\bigg(\bigcap_{\tau=a_\ell}^{a_{\ell+1}-1} A_\tau \bigg| J_{m_\ell - t_0}\bigg) \\
&= \mathbb{P}\left(M_{m_\ell} \ge (b_{m_\ell} f_{m_\ell})^{k+t_0-m_\ell} \text{ for all } a_\ell \le k < a_{\ell+1} - 1 | F_{m_\ell} \ge f_{m_\ell}/(1-\beta)\right) \\
&\ge \mathbb{P}\left(M_{m_\ell} \ge (b_{m_\ell} f_{m_\ell})^{k+t_0-m_\ell} \text{ for all } k \ge 0 | F_{m_\ell} \ge f_{m_\ell}/(1-\beta)\right) \ge 1 - \xi_\ell,
\end{aligned}$$

where we have used (20) with $f \mapsto f_{m_\ell}$. Therefore,

$$\mathbb{P}(C_t) \ge \left(\prod_{\tau=1}^{t} (1-\eta_\tau)\right)\left(\prod_{\ell=0}^{\ell'} (1-\xi_\ell)\right) \ge 1 - \sum_{\tau=1}^{t} \eta_\tau - \sum_{\ell=0}^{\ell'} \xi_\ell.$$

By (42) and (43), we have

$$\lim_{t_0\to\infty} \mathbb{P}(C) = 1.$$

Since $C \subset E$ and $C \subset J$, the proof is completed. □

Lemma 5.7 (Lower bound). *Almost surely on survival,*

$$\liminf_{t\to\infty} \frac{\log X(t)}{t \log^{(n)}(t)} \ge \frac{1}{\alpha}, \quad \liminf_{t\to\infty} \frac{W_t}{u_n(t)} \ge 1.$$

In other words,

$$\mathbb{P}\left(\liminf_{t\to\infty} \frac{\log X(t)}{t \log^{(n)}(t)} \ge \frac{1}{\alpha}\right) = \mathbb{P}\left(\liminf_{t\to\infty} \frac{W_t}{u_n(t)} \ge 1\right) = \mathbb{P}(\mathfrak{A}) = p_s.$$

Proof. Fix ε and ε_1 as in Lemma 5.2. For any $0 < \varepsilon'$, Lemma 5.6 implies the existence of t_0 such that

$$\mathbb{P}\left(\log \mathcal{Y}(t) \ge \frac{1-\varepsilon}{\alpha}(t+t_0)\log^{(n)}(t+t_0) \text{ for all } t \ge 0\right) \ge 1 - \varepsilon',$$

$$\mathbb{P}\left(W_t \ge (1-\varepsilon_1)u_n(t+t_0) \text{ for all } t \ge 0\right) \ge 1 - \varepsilon'.$$

Since W_t as well as $X(t)$ is unbounded on survival (Lemma 4.1), there should be τ and $k \ge 1$ almost surely on survival such that $W_\tau > (1-\varepsilon_1)u_n(t_0)$ and $N > \mathcal{Y}(0)$, where N is the number of individual with fitness W_τ at generation $\tau + k$. Now couple $X(t+\tau+k)$ with $\mathcal{Y}(t)$, which gives $X(t+\tau+k) \ge \mathcal{Y}(t)$ for all $t \ge 0$. We denote the event that has such τ and k by D. Note that $\mathbb{P}(D \cap \mathfrak{A}) = p_s$ by Lemma 4.1 and, obviously, $\mathbb{P}(D) \ge p_s$. Therefore,

$$\begin{aligned}
p_s &\ge \mathbb{P}\left(\liminf_{t\to\infty} \frac{\log X(t)}{t \log^{(n)}(t)} \ge \frac{1-\varepsilon}{\alpha}\right) \\
&\ge \mathbb{P}\left(\liminf_{t\to\infty} \frac{\log X(t)}{t \log^{(n)}(t)} \ge \frac{1-\varepsilon}{\alpha} \middle| D\right) \mathbb{P}(D) \\
&\ge \mathbb{P}\left(\log \mathcal{Y}(t) \ge \frac{1-\varepsilon}{\alpha}(t+t_0)\log^{(n)}(t+t_0) \text{ for all } t \ge 0\right) \mathbb{P}(D) \ge (1-\varepsilon')p_s,
\end{aligned}$$

where we have used the Markov property. By the same token, we have

$$p_s \geq \mathbb{P}\left(\liminf_{t\to\infty} \frac{W_t}{u_n(t)} \geq 1-\varepsilon_1\right) \geq (1-\varepsilon')p_s.$$

Since ε' and ε are arbitrary, the proof is completed. □

By Lemma 5.4 and Lemma 5.7, Theorem 1 is proved.

6 Proof of Theorem 2

This section presents two lemmas, which will prove Theorem 2. Needless to say, G is always of type II throughout this section. For convenience, we define

$$\chi(t,n,\nu) := \begin{cases} t^\nu, & n=1, \\ \exp(t^\nu), & n=2, \\ \exp\left(t\left(\log^{(n-2)}(t)\right)^{-\nu}\right), & n\geq 3, \end{cases}$$

$$U(t,n,\nu,a) := \chi(t,n,\nu)\left(\log^{(\max\{0,n-2\})}(t)\right)^{-a}, \quad \mathcal{G}(x,n,a) := \log x\left(\log^{(n)}(x)\right)^{a},$$

with an appropriate domain. Again, the distinction between the MMM and the FMM does not play any role in the proof of Theorem 2.

Lemma 6.1 (**Variation of Lemma 5.3**). *Assume $X(0)<\infty$ and $Q_0<\infty$, fix $\varepsilon>0$ and let*

$$\nu_n := \begin{cases} (1+2\varepsilon)(1+\alpha)/\alpha, & n=1, \\ \varepsilon+1/(1+\alpha), & n=2, \\ \alpha/(1+\varepsilon)^2, & n\geq 3, \end{cases} \quad a_n := \begin{cases} 1+\varepsilon, & n=1, \\ \alpha/(1+\alpha), & n=2, \\ \alpha/(1+\varepsilon), & n\geq 3. \end{cases}$$

Then

$$\lim_{m\to\infty} \mathbb{P}\left(\log\Xi(t) \leq \chi(t+m,n,\nu_n),\ \log W_t \leq U(t+m,n,\nu_n,a_n) \text{ for all } t\right) = 1.$$

Proof. We first make a precise criterion as to the meaning of large m. Obviously, there is m_1 such that $\chi(m,n,\nu_n)\geq \log X(0)$ and $U(m,n,\nu_n,a_n)\geq \log Q_0$ for all $m>m_1$. Let $H(x) := \chi(x+1,n,\nu_n)-\chi(x,n,\nu_n)-U(x,n,\nu_n,a_n)$. By the mean value theorem, there is x_0 ($x\leq x_0\leq x+1$) such that

$$\begin{aligned} \chi(x+1,n,\nu_n)-\chi(x,n,\nu_n) &= \left.\frac{\partial\chi(x,n,\nu_n)}{\partial x}\right|_{x=x_0} \\ &= U(x_0,n,\nu_n,a_n)\times\begin{cases} \nu_n x_0^{\varepsilon}, & n\leq 2, \\ \left(\log^{(n-2)}(x_0)\right)^{\varepsilon\nu_n}\left(1-\nu_n\prod_{k=1}^{n-2}\frac{1}{\log^{(k)}(x_0)}\right), & n\geq 3, \end{cases} \end{aligned}$$

which gives $\lim_{x\to\infty}H(x)=\infty$. Therefore, there is m_2 such that $H(x)>2$ for all $x>m_2$. Let $\varepsilon_0=\varepsilon/(1+\varepsilon)$. By definition, there is m_3 such that $\log G(x)\leq -\mathcal{G}(x,n,\alpha/(1+\varepsilon_0))$ for all $x>m_3$. Since $(\nu_1-a_1)\alpha > a_1(1+\varepsilon_0)$, $\nu_2\alpha > a_2(1+\varepsilon_0)$, and $\alpha > a_n(1+\varepsilon_0)$ for $n\geq 3$, we have

$$\begin{aligned} &\lim_{t\to\infty}\frac{\mathcal{G}\left(\exp\left(U(t,n,\nu_n,a_n)\right),n,\alpha/(1+\varepsilon_0)\right)}{\chi(t,n,\nu_n)} \\ &= \lim_{t\to\infty}\left(\log^{(\max\{0,n-2\})}(t)\right)^{-a_n}\left(\log^{(n-1)}\left(U(t,n,\nu_n,a_n)\right)\right)^{\alpha/(1+\varepsilon_0)} = \infty. \end{aligned}$$

Therefore, there is m_4 such that $\mathcal{G}\left(\exp\left(U(t,n,\nu_n,a_n)\right),n,\alpha/(1+\varepsilon_0)\right) \geq 2\chi(t,n,\nu_n)$ and, accordingly,

$$G\left(\exp\left(U(t,n,\nu_n,a_n)\right)\right) \leq e^{-2\chi(t,n,\nu_n)}, \tag{45}$$

for all $t > \max\{m_3, m_4\}$. We set $m_0 = \max\{m_1,m_2,m_3,m_4\}$ and we assume $m > m_0$ in what follows. For given m, we define $\tau_t := t+m$ and

$$E_t := \{\log W_t \leq U(\tau_t,n,\nu_n,a_n)\}, \quad E'_t := \{\log Q_t \leq U(\tau_t,n,\nu_n,a_n)\},$$
$$A_t := \{\log \Xi(t) \leq \chi(\tau_t,n,\nu_n)\}, \quad A = \bigcap_{k=1}^{\infty} A_k.$$

We can repeat (36) for E_t and E'_t.

$$\bigcap_{k=0}^{t} E_k = \bigcap_{k=0}^{t} E'_k. \tag{46}$$

We also define, for $t \geq 1$, $D_0 = A_0 \cap E_0$, $D_t = A_t \cap E_t \cap D_{t-1}$, and $D = \bigcap_{k=1}^{\infty} D_k$. Observe that $\mathbb{P}(D_t) = \mathbb{P}(E_t|A_t \cap D_{t-1})\mathbb{P}(A_t|D_{t-1})\mathbb{P}(D_{t-1})$. Using Lemma 4.5 and (46), we have

$$\mathbb{P}(A_k|D_{k-1}) = \mathbb{P}(A_k|A_{k-1} \cap E'_{k-1}) \geq 1 - 2\exp\left(-e^{U(\tau_{k-1},n,\nu_n,a_n)+\chi(\tau_{k-1},n,\nu_n)}\right) =: 1-\xi_k,$$

where we have used (21) with $f \mapsto e^{U(\tau_{k-1},n,\nu_n,a_n)}$, $x \mapsto e^{\chi(\tau_{k-1},n,\nu_n)}$, and $B \mapsto e^{H(\tau_{k-1})} \geq e^2$. Since W_t is purely determined by $\Xi(t)$, we have

$$\mathbb{P}(E_t|A_t \cap D_{t-1}) = \mathbb{P}(E_t|A_t) \geq 1 - \beta\exp\left(-\chi(\tau_k,n,\nu_n,a_n)\right) =: 1-\eta_k,$$

where we have used (22) with $y \mapsto e^{\chi(\tau_k,n,\nu_n)}$, $x \mapsto e^{U(\tau_k,n,\nu_n,a_n)}$, and (45). Therefore, we have

$$\mathbb{P}(D) \geq \prod_{k=1}^{\infty}(1-\xi_k)(1-\eta_k) \geq 1 - \sum_{k=1}^{\infty}(\xi_k+\eta_k). \tag{47}$$

Since $\lim_{k\to\infty}(\xi_k+\eta_k)\tau_k^2 = 0$ and $\tau_k^{-2} < k^{-2}$, the series in (47) converges uniformly for large m. Since $\lim_{m\to\infty}(\xi_k+\eta_k) = 0$ for all k, we have $\lim_{m\to\infty}\mathbb{P}(D) = 1$, which completes the proof. $\square$

Definition (Initial condition for Lemma 6.3). Fix $0 < \varepsilon < 1/\alpha$ and let

$$\nu_n := \begin{cases} (1+\alpha)/[\alpha(1+2\varepsilon)], & n=1, \\ 1/[(1+\alpha)(1+2\varepsilon)], & n=2, \\ \alpha(1+3\varepsilon), & n\geq 3, \end{cases} \quad a_n := \begin{cases} (1+\alpha)/(1+\alpha+\alpha\varepsilon), & n=1, \\ \alpha/(1+\alpha), & n=2, \\ \alpha(1+2\varepsilon), & n\geq 3, \end{cases}$$

which should not be confused with ν_n and a_n defined in Lemma 6.1. Note that $\nu_1 > a_1$ because $\varepsilon < 1/\alpha$. Define

$$f_k := (1-\beta)\exp\left(U(k,n,\nu_n,a_n)\right), \quad b_k := \frac{1}{1-\beta}\exp\left(-\frac{\varepsilon}{1+\varepsilon}U(k,n,\nu_n,a_n)\right),$$
$$f_k b_k = \exp\left(\frac{U(k,n,\nu_n,a_n)}{1+\varepsilon}\right).$$

Once k_0 is determined as in Lemma 6.2, we define the initial condition with an integer t_0 larger than k_0 in exactly the same way as in the previous section. We use $M_k(t)$, F_k, F_{k+t_0}, and $\mathcal{Y}(t)$ with an appropriate modification of the meaning.

Lemma 6.2. *For ε, ν_n, a_n, b_k, f_k defined above there is an integer k_0, which is larger than* $\exp^{(n)}(0)$*, such that for all $m \geq k_0$ we have*

(Condition 1) $0 < b_m < b_c$, $(1 - b_m + b_m \log b_m) f_m > 1$, *and* $b_m f_m > e$*;*

(Condition 2) $G\left(\exp(U(m, n, \nu_n, a_n))\right) \geq \exp\left(-\frac{1}{2}\chi(m, n, \nu_n)\right)$.

Proof. Obviously, there is an integer k_1 that satisfies (**Condition 1**) for all $m \geq k_0$. By definition, we have $\log G(x) \geq -\mathcal{G}(x, n, \alpha(1+\varepsilon))$ for all sufficiently large x. Since $(\nu_1 - a_1)\alpha(1+\varepsilon) < a_1$, $\nu_2 \alpha(1+\varepsilon) < a_2$, and $a_n > \alpha(1+\varepsilon)$ for $n \geq 3$, we have

$$\begin{aligned} &\lim_{y\to\infty} \frac{\mathcal{G}(\exp(U(y, n, \nu_n, a_n)))}{\chi(y, n, \nu_n)} \\ &= \lim_{y\to\infty} \left(\log^{(\max\{0,n-2\})}(y)\right)^{-a_n} \left(\log^{(n-1)}\left(U(y, n, \nu_n, a_n)\right)\right)^{\alpha(1+\varepsilon)} = 0, \end{aligned}$$

which guarantees the existence of an integer k_2 such that

$$\log G\left(\exp(U(y, n, \nu_n, a_n))\right) \geq -\frac{1}{2}\chi(y, n, \nu_n) \tag{48}$$

for all $y \geq k_2$. Now we set $k_0 := \max\{k_1, k_2\}$, which completes the proof. □

Lemma 6.3 (**Variation of Lemma 5.6**). *For the initial conditions defined above with $t_0 \geq k_0$, we define two events*

$$E_t := \{\log \mathcal{Y}(t) \geq \chi(t + t_0, n, \nu_n)\}, \quad E := \bigcap_{t=1}^{\infty} E_t,$$

$$J_t := \{\log W_t \geq U(t + t_0, n, \nu_n, a_n)\}, \quad J := \bigcap_{t=1}^{\infty} J_t.$$

Then,

$$\lim_{t_0\to\infty} \mathbb{P}(E) = \lim_{t_0\to\infty} \mathbb{P}(J) = 1.$$

Proof. Let

$$m_t := \begin{cases} \left\lfloor \frac{1}{2}(t + t_0) \right\rfloor, & n = 1, \\ \left\lfloor t + t_0 - \frac{1}{2}(t + t_0)^{1-\nu_2} \right\rfloor, & n = 2, \\ \left\lfloor t + t_0 - \frac{1}{2}\left(\log^{(n-2)}(t + t_0)\right)^{\nu_n} \right\rfloor, & n \geq 3. \end{cases}$$

Assume t_0 is so large that $m_0 > k_0$ and $(m_t)_{t\geq 0}$ is an non-dereasing sequence of t. Since $1 > a_1$, $1 > \nu_2 + a_2$, and $\nu_n > a_n$ for $n \geq 3$, we have

$$\lim_{t_0\to\infty} \frac{\chi(t + t_0, n, \nu_n)}{(t + t_0 - m_t) U(m_t, n, \nu_n, a_n)} = \lim_{t\to\infty} \frac{\chi(t + t_0, n, \nu_n)}{(t + t_0 - m_t) U(m_t, n, \nu_n, a_n)} = 0.$$

So there is t_1 such that $(t + t_0 - m_t)U(m_t, n, \nu_n, a_n) \geq (1+\varepsilon)\chi(t + t_0, n, \nu_n)$ for all $t_0 > t_1$ and $t \geq 0$. In the following, we assume $t_0 > t_1$. Define

$$A_t := \left\{ M_{m_t}(t) \geq (b_{m_t} f_{m_t})^{t+t_0-m_t} \right\}, \quad \tilde{J} := \bigcap_{k=m_0}^{t_0} \{F_k \geq f_k/(1-\beta)\},$$

$$C_0 = A_0 \cap \tilde{J}, \quad C_t = A_t \cap J_t \cap C_{k-1}, \quad C = \bigcap_{t=1}^{\infty} C_t.$$

Note that $\tilde{J}$ and A_0 are sure events (by the initial condition) and so is C_0. Also note that $A_t \subset E_t$. Observe that $\mathbb{P}(C_t) = \mathbb{P}(J_t|A_t \cap C_{t-1})\,\mathbb{P}(A_t|C_{t-1})\,\mathbb{P}(C_{t-1})$. Since W_t is solely determined by $\Xi(t)$ and $\log \Xi(t) \geq \chi(t+t_0, n, \nu_n)$ on the event $A_t \cap C_{t-1}$, we have

$$\mathbb{P}(J_t|A_t \cap C_{t-1}) \geq 1 - \exp\left(-\beta e^{\chi(t+t_0,n,\nu_n)/2}\right) =: 1 - \eta_t,$$

where we have used (23) with $y \mapsto \exp(\chi(t+t_0, n, \nu_n))$, $x \mapsto \exp(U(m_t, n, \nu_n, a_n))$, (48), and $\chi(m_t, n, \nu_n) \leq \chi(t+t_0, n, \nu_n)$. Therefore, we have

$$\mathbb{P}(C) \geq \left(\prod_{\tau=1}^{\infty}(1-\eta_\tau)\right)\prod_{\tau=1}^{\infty}\mathbb{P}(A_\tau|C_{\tau-1}).$$

Note that A_τ is independent of J_k for $m_\tau < k < \tau$ and of A_k for $k < a_\ell$ (this is because $M_m(t)$'s for different m's are mutually independent branching processes). Since $m_{t+1} - m_t \leq 1$, all F_k with $k \geq m_0$ should affect a certain A_τ at least once. Therefore,

$$\begin{aligned}\prod_{\tau=1}^{\infty}\mathbb{P}(A_\tau|C_{\tau-1}) &\geq \prod_{\tau=1}^{\infty}\mathbb{P}\left(M_{m_\tau}(k) \geq (b_{m_\tau}f_{m_\tau})^{k+t_0-m_\tau} \text{ for all } k \geq 0 | F_{m_\tau} \geq f_{m_\tau}/(1-\beta)\right)\\ &\geq \prod_{\tau=1}^{\infty}(1-\xi_\tau) \geq 1 - \sum_{\tau=1}^{\infty}\xi_\tau,\end{aligned}$$

where we have used (20) with $f \mapsto f_{m_\tau}$. Therefore,

$$\mathbb{P}(C) \geq 1 - \sum_{\tau=1}^{t}(\eta_\tau + \xi_\tau)\,. \tag{49}$$

Since $\lim_{k\to\infty}(\xi_k + \eta_k)(k+t_0)^2 = 0$, there is a constant c that is independent of t_0 such that $\xi_k + \eta_k \leq c(k+t_0)^{-2} \leq ck^{-2}$ for all k. Therefore, the series in (49) converges uniformly. Since $\lim_{t_0\to\infty}(\xi_k + \eta_k) = 0$ for any k, we have $\lim_{t_0\to\infty}\mathbb{P}(C) = 1$. Since $C \subset E$ and $C \subset J$, we get the desired result. □

By the same logic as in Lemma 5.4 and Lemma 5.7, Lemma 6.1 and Lemma 6.3 now prove Theorem 2.

7 The empirical fitness distribution

In this section, we introduce two variants of the FMM that (completely or partially) neglect fluctuations in the original model with the type I tail function. These variants will be called the deterministic FMM (DFMM) and semi-deterministic FMM (SFMM) and will be defined in Section 7.2 and Section 7.3, respectively. As we will see presently, neglecting some fluctuations will facilitate rigorous proofs for the limit behaviour of the EFD.

To explain the motivation of introducing the DFMM and SFMM, we begin by finding in Lemma 7.3 tighter bounds for $\mathcal{X}_t$ of the Galton-Watson process, which show that the fluctuations of $N_k(t)$ become smaller and smaller over time.

7.1 Fluctuations of $N_k(t)$ and W_t

Lemma 7.1. *If $B > 1$, $\theta > 0$, and $1 \le x_1 < x_2 - 1$, then*

$$\mathbb{P}(\mathcal{X}_t \le Bx_2\theta | x_1 \le \mathcal{X}_{t-1} \le x_2) \ge 1 - \frac{B}{B-1} e^{-x_1\theta(B\log B - B + 1)}.$$

Proof. Let $m \ge x_1$ and $B' = Bx_2/m \ge B$. By Lemma 4.3 together with Remark 4.1, we have

$$\begin{aligned}\mathbb{P}(\mathcal{X}_t > Bx_2\theta|\mathcal{X}_{t-1} = m) &= \mathbb{P}(\mathcal{X}_t > B'm\theta|\mathcal{X}_{t-1} = m) \\ &\le \tfrac{B'}{B'-1} e^{-m\theta(B'\log B' - B' + 1)} \le \tfrac{B}{B-1} e^{-m\theta(B\log B - B + 1)},\end{aligned}$$

where we have used the fact that $y/(y-1)$ and $y(1-\log y)$ are decreasing functions in the region $y > 1$. Since $m \ge x_1$, we have the desired result. ☐

Lemma 7.2. *If $1 < B < \frac{3}{2}$, $0 < b < 1$, $(1 - b + b\log b)\theta > 1$, and $1 \le x_1 < x_2 - 1$, then*

$$\mathbb{P}(bx_1\theta \le \mathcal{X}_t \le Bx_2\theta | x_1 \le \mathcal{X}_{t-1} \le x_2) \ge 1 - x_1\theta e^{-x_1\theta(1-b)^2/2} - \frac{B}{B-1} e^{-x_1\theta(B-1)^2/3}.$$

Proof. Using Lemma 4.4 with $f = \theta$ and Lemma 7.1, we have

$$\mathbb{P}(bx_1\theta \le \mathcal{X}_t \le Bx_2\theta | x_1 \le \mathcal{X}_{t-1} \le x_2) \ge 1 - x_1\theta e^{-x_1\theta(1-b+b\log b)} - \frac{B}{B-1} e^{-x_1\theta(1-B+B\log B)}.$$

Since $1 - x + x\log x \ge \begin{cases} \frac{1}{2}(1-x)^2, & 0 < x < 1, \\ \frac{1}{3}(x-1)^2, & 1 < x < 3/2, \end{cases}$ the proof is completed. ☐

Lemma 7.3. *Fix $0 < \varepsilon < \frac{1}{2}$ and abbreviate $c := (1-\varepsilon)/2$. Let $a_t := \theta^{-(1-2\varepsilon)/2}\left(1 - \theta^{-ct}\right)$ and*

$$\begin{aligned} b_t &:= \frac{1 - a_t}{1 - a_{t-1}} = 1 - \frac{\theta^c - 1}{1 - a_{t-1}} \theta^{-ct-(1-2\varepsilon)/2}, \quad \prod_{k=1}^{t} b_k = 1 - a_t. \\ B_t &:= \frac{1 + a_t}{1 + a_{t-1}} = 1 + \frac{\theta^c - 1}{1 + a_{t-1}} \theta^{-ct-(1-2\varepsilon)/2}, \quad \prod_{k=1}^{t} B_k = 1 + a_t, \end{aligned}$$

where θ is assumed so large that for all $t \ge 1$

$$\begin{aligned} 0 < a_t < 1, \quad 1 < B_t < \frac{3}{2}, \quad \theta^t \ge \theta^{ct-(1-2\varepsilon)/2}, \\ \frac{(1-\theta^{-c})^2}{2(1-a_{t-1})} \ge \frac{1}{4}, \quad \frac{(1-a_{t-1})(1-\theta^{-c})^2}{3(1+a_{t-1})} \ge \frac{1}{4}, \quad \frac{B_t(1+a_{t-1})}{\theta^c - 1} \le 1, \\ 4\theta \exp\left(-\frac{\theta^{2\varepsilon}}{4}\right) \le \exp\left(-\frac{\theta^{2\varepsilon}}{5}\right), \quad \theta^t \exp\left(-\frac{1}{4}\theta^{\varepsilon(t+1)}\right) \le \frac{12}{\pi^2 t^2} \theta \exp\left(-\frac{\theta^{2\varepsilon}}{4}\right). \end{aligned} \tag{50}$$

Then,

$$\mathbb{P}\left(\left|\frac{\mathcal{X}_t}{\theta^t} - 1\right| \le 2\theta^{-(1-2\varepsilon)/2} \text{ for all } t\right) \ge 1 - \exp\left(-\frac{\theta^{2\varepsilon}}{5}\right). \tag{51}$$

Proof. Let $A_t := \{(1-a_t)\theta^t \le \mathcal{X}_t \le (1+a_t)\theta^t\}$ and $A := \bigcap_{t=0}^{\infty} A_t$. Abbreviating $x_1 := (1-a_{t-1})\theta^{t-1}$ and $x_2 := (1+a_{t-1})\theta^{t-1}$, we can write

$$\mathbb{P}(A_t|A_{t-1}) = \mathbb{P}\left(b_t x_1 \theta \le \mathcal{X}_t \le B_t x_2 \theta | x_1 \le \mathcal{X}_{t-1} \le x_2\right).$$

By Lemma 7.2, we have

$$\begin{aligned}\mathbb{P}(A_t|A_{t-1}) \ge 1 &- (1-a_{t-1})\theta^t \exp\left(-\frac{(1-\theta^{-c})^2}{2(1-a_{t-1})}\theta^{\varepsilon(t+1)}\right)\\ &- \frac{B_t(1+a_{t-1})}{\theta^c - 1}\theta^{ct-(1-2\varepsilon)/2}\exp\left(-\frac{(1-a_{t-1})(1-\theta^{-c})^2}{3(1+a_{t-1})}\theta^{\varepsilon(t+1)}\right)\\ &\ge 1 - 2\theta^t \exp\left(-\frac{1}{4}\theta^{\varepsilon(t+1)}\right),\end{aligned}$$

where we have used (50). Using the last condition of (50), we have

$$\sum_{t=1}^{\infty} \theta^t \exp\left(-\frac{1}{4}\theta^{\varepsilon(t+1)}\right) \le \frac{12}{\pi^2}\theta \exp\left(-\frac{\theta^{2\varepsilon}}{4}\right)\sum_{t=1}^{\infty}\frac{1}{t^2} = 2\theta \exp\left(-\frac{\theta^{2\varepsilon}}{4}\right).$$

Hence,

$$\mathbb{P}(A) \ge 1 - 2\sum_{t=1}^{\infty}\theta^t \exp\left(-\frac{\theta^{\varepsilon(t+1)}}{4}\right) \ge 1 - 4\theta\exp\left(-\frac{\theta^{2\varepsilon}}{4}\right) \ge 1 - \exp\left(-\frac{\theta^{2\varepsilon}}{5}\right),$$

where we have used Lemma 4.5. Since $1-\theta^{-ct} \le 2$ and, therefore, $a_t \le 2\theta^{-(1-2\varepsilon)/2}$, the probability in (51) is larger than $\mathbb{P}(A)$ and the proof is completed. □

Definition. By $\theta_0(\varepsilon)$ we denote the infimum over all θ that satisfy (50).

Remark 7.1. *If we are given a weaker condition in Lemma 7.3 such that there are x and y such that $x \ge \theta \ge y > \theta_0(\varepsilon)$, then we have*

$$\mathbb{P}\left(\left|\frac{\mathcal{X}_t}{\theta^t} - 1\right| \le 2\theta^{-(1-2\varepsilon)/2} \text{ for all } t \,\middle|\, y \le \theta \le x\right) \ge 1 - \exp\left(-\frac{y^{2\varepsilon}}{5}\right).$$

Lemma 7.4. *For a discrete-time stochastic process Z_t and a nonzero function $f(t)$, define*

$$J := \left\{\lim_{t\to\infty}\frac{Z_t}{f(t)} = 1\right\}, \quad D_{m,k} := \left\{\left|\frac{Z_k}{f(k)} - 1\right| \le 2^{-m}\right\},$$

$$O_{m,\tau} := \bigcap_{k=\tau}^{\infty} D_{m,k}, \quad O_m := \bigcup_{\tau=1}^{\infty} O_{m,\tau}, \quad O = \bigcap_{m=1}^{\infty} O_m.$$

Then, $O = J$ and

$$\lim_{\tau\to\infty}\mathbb{P}(O_{m,\tau}) \ge \mathbb{P}(J),$$

for any positive integer m.

Proof. First note that $O_{m,\tau} \subset O_{m,\tau+1} \subset O_m$ and $O \subset O_{m+1} \subset O_m$. Consider any outcome $\omega \in J$ and fix m. Under ω, for any $0 < \varepsilon' \le 2^{-m}$ there is k_0 such that $|Z_k/f(k) - 1| \le \varepsilon' \le 2^{-m}$ for all $k \ge k_0$, which implies $\omega \in O_m$. Since m is arbitrary, we have $J \subset O$.

Now consider $\omega' \notin J$. Then under ω' there is $\varepsilon' > 0$ such that $|Z_k/f(k) - 1| > \varepsilon'$ for infinitely many k's. Hence, ω' cannot be an outcome in O_m if $2^{-m} < \varepsilon'$. Hence, $\omega' \notin O$ and, accordingly, $O \subset J$. Even if J is empty, the proof of $O \subset J$ is still applicable and the rest of the statement is trivially valid.

Since $O \subset O_m$ and $\mathbb{P}(O_m) = \lim_{\tau\to\infty}\mathbb{P}(O_{m,\tau})$ for any m, the proof is completed. □

Lemma 7.5. *If G is of the type I with $n = 1$ or with $n = 2$ and $\alpha < 1$, then let*

$$J := \left\{ \lim_{t\to\infty} \frac{W_t}{u_n(t)} = 1 \right\}. \tag{52}$$

If G is of the type II, then let

$$J := \left\{ \lim_{t\to\infty} \frac{\log^{(2)}(W_t)}{\log t} = \frac{1}{\alpha} \right\}$$

for $n = 1$,

$$J := \left\{ \lim_{t\to\infty} \frac{\log^{(3)}(W_t)}{\log t} = \frac{1}{1+\alpha} \right\}$$

for $n = 2$, and

$$J := \left\{ \lim_{t\to\infty} \frac{1}{\log^{(n-1)}(t)} \log\left(\frac{\log^{(2)}(W_t)}{t} \right) = -\alpha \right\}$$

for $n \geq 3$. For the type II tail function or for the type I tail function with $n = 1$, fix an arbitrary ε satisfying $0 < \varepsilon < 1/2$. For the type I tail function with $n = 2$ and $\alpha < 1$, fix an arbitrary ε satisfying $\alpha < 2\varepsilon < 1$. Abbreviate $\theta_k := (1-\beta)W_k$ and let

$$C_k := \bigcap_{t=k}^{\infty} \left\{ \left| \frac{N_k(t)}{\theta_k^{t-k}} - 1 \right| \leq 2\theta_k^{-(1-2\varepsilon)/2} \right\}, \quad E_\tau := \bigcap_{k=\tau}^{\infty} C_k, \quad E := \bigcup_{\tau=1}^{\infty} E_\tau,$$

where we assume $N_k(t)/\theta_k^{t-k} = 1$ and $\theta_k^{-(1-2\varepsilon)/2} = \infty$ if $W_k = 0$. Then $\mathbb{P}(E|J) = 1$.

Proof. Let $U(x,m) := (1-2^{-m})u_n(x)$ for the type I tail function and

$$U(x,m) := \exp\left(x^{(1-2^{-m})/\alpha} \right),$$
$$U(x,m) := \exp^{(2)}\left(x^{(1-2^{-m})/(1+\alpha)} \right),$$
$$U(x,m) := \exp^{(2)}\left(x \exp\left(-\alpha(1+2^{-m}) \log^{(n-1)}(x) \right) \right),$$

for the type II tail function with $n = 1$, $n = 2$, and $n \geq 3$, respectively. In the above definition, x is assumed sufficiently large that $U(x,m)$ is well defined. With the fixed ε, for any positive m there is $\tau_0(m)$ such that

$$\exp\left(-\frac{(1-\beta)^{2\varepsilon}}{5} U(t,m)^{2\varepsilon} \right) \leq \frac{1}{t(t+1)}, \tag{53}$$

for all $t \geq \tau_0(m)$. Let

$$D_{m,k} := \left\{ \left| \frac{W_k}{u_n(k)} - 1 \right| \leq 2^{-m} \right\},$$

for the type I tail function and

$$D_{m,k} := \left\{ \left| \frac{\alpha \log^{(2)}(W_k)}{\log k} - 1 \right| \leq 2^{-m} \right\},$$
$$D_{m,k} := \left\{ \left| \frac{(1+\alpha)\log^{(3)}(W_k)}{\log k} - 1 \right| \leq 2^{-m} \right\},$$
$$D_{m,k} := \left\{ \left| \frac{1}{\alpha \log^{(n-1)}(k)} \log\left(\frac{\log^{(2)}(W_k)}{\log k} \right) + 1 \right| \leq 2^{-m} \right\},$$

for the type II tail function with $n = 1$, $n = 2$, and $n \geq 3$, respectively. Let

$$C_k^c := \bigcup_{t=k}^{\infty} \left\{ \left| \frac{N_k(t)}{\theta_k^{t-k}} - 1 \right| > 2\theta_k^{-(1-2\varepsilon)/2} \right\}, \quad E_\tau^c := \bigcup_{k=\tau}^{\infty} C_k^c,$$

$$O_{m,\tau} := \bigcap_{k=\tau}^{\infty} D_{m,k}, \quad O_m := \bigcup_{\tau=1}^{\infty} O_{m,\tau},$$

where τ is assumed large enough such that $u_n(\tau)$, $\log \tau$, and $\log^{(n-1)}(\tau)$ are well defined. Note that, for all sufficiently large τ, $E_\tau \subset E_{\tau+1} \subset E$. Now, consider $\mathbb{P}(E_\tau^c \cap O_{m,\tau})$ for $m \geq 1$. By the sub-additivity of the probability measure, we have

$$\mathbb{P}(E_\tau^c \cap O_{m,\tau}) = \mathbb{P}\left(\bigcup_{k=\tau}^{\infty} (C_k^c \cap O_{m,\tau}) \right) \leq \sum_{k=\tau}^{\infty} \mathbb{P}\left(C_k^c \cap O_{m,\tau}\right) \leq \sum_{k=\tau}^{\infty} \mathbb{P}(C_k^c \cap D_{m,k}),$$

where we have used $O_{m,\tau} \subset D_{m,k}$ for any $k \geq \tau$. Now fix an integer $m \geq 1$ and consider large enough τ such that $\tau > \tau_0(m)$ as in (53) and $(1-\beta)U(k,m) > \theta_0(\varepsilon)$ for all $k \geq \tau$. Since $\mathbb{P}(C_k^c \cap D_{m,k}) \leq \mathbb{P}(C_k^c | D_{m,k}) = 1 - \mathbb{P}(C_k | D_{m,k})$, Remark 7.1 with $y \mapsto (1-\beta)U(k,m)$ gives

$$\mathbb{P}(C_{k+\tau}^c \cap D_{m,k+\tau}) \leq \frac{1}{(k+\tau)(k+\tau+1)},$$

for all $k \geq 0$, where we have used (53). Therefore, we have

$$\lim_{\tau\to\infty} \mathbb{P}(E_\tau^c \cap O_{m,\tau}) \leq \lim_{\tau\to\infty} \sum_{k=0}^{\infty} \mathbb{P}(C_{\tau+k}^c \cap D_{m,\tau+k}) \leq \lim_{\tau\to\infty} \frac{1}{\tau} = 0. \tag{54}$$

Since $\mathbb{P}(O_{m,\tau}) \leq p_s$ for all sufficiently large τ, $\mathbb{P}(J) = p_s$, and $\mathbb{P}(E_\tau \cap O_{m,\tau}) = \mathbb{P}(O_{m,\tau}) - \mathbb{P}(E_\tau^c \cap O_{m,\tau})$, Lemma 7.4 gives

$$\lim_{\tau\to\infty} \mathbb{P}(E_\tau \cap O_{m,\tau}) = \lim_{\tau\to\infty} \mathbb{P}(O_{m,\tau}) = p_s. \tag{55}$$

Since $E_\tau \subset E$ and $O_{m,\tau} \subset O_m$, we have $p_s \geq \mathbb{P}(E \cap O_m) \geq \lim_{\tau\to\infty} \mathbb{P}(E_\tau \cap O_{m,\tau}) = p_s$, for all m. Therefore, $\mathbb{P}(E|J)\mathbb{P}(J) = \mathbb{P}(E \cap J) = \lim_{m\to\infty} \mathbb{P}(E \cap O_m) = p_s$. Since $\mathbb{P}(J) = p_s$, the proof is completed. □

Remark 7.2. *In the proof,* (53) *plays the decisive role. If G is of type I with $n \geq 3$ or with $n = 2$ and $\alpha > 1$,* (53) *is not applicable. Within the tools we are equipped with, we are not aware of a similar result to Lemma 7.5 for fast decaying tail functions.*

Remark 7.3. *We can rewrite Lemma 7.5 as follows. For any type II tail function and for a type I tail function with $n = 1$ or with $n = 2$ and $\alpha < 1$, for any $\varepsilon > 0$*

$$\lim_{\tau\to\infty} \mathbb{P}\left(\left| \frac{N_k(k+s)}{(1-\beta)^s W_k^s} - 1 \right| < \varepsilon \text{ for all } s \geq 0 \text{ and for all } k \geq \tau \right) = p_s.$$

Remark 7.4. *If G is of the Fréchet type in [1], then setting*

$$J := \left\{ \lim_{t\to\infty} \frac{\log^{(2)}(W_t)}{t} = \nu(\alpha) \right\}$$

in Lemma 7.5 gives $\mathbb{P}(E|J) = 1$.

The following two lemmas, which will not be directly used later, are for explaining at what point the proof of Theorem 3 becomes difficult and also for providing a more compelling rationale of introducing DFMM and SFMM.

Lemma 7.6. *For the FMM, define two random sequences* (η_t) *and* (ξ_t) *as*

$$X(t) = (1-\beta)\Xi(t) + \eta_t \Xi(t)^{3/4}, \quad (1-\beta)\Xi(t) = X(t) + \xi_t X(t)^{3/4}.$$

In case $X(t) = \Xi(t) = 0$*, we define* $\xi_t = \eta_t = 0$*. Then almost surely*

$$\lim_{t\to\infty} \eta_t = \lim_{t\to\infty} \xi_t = 0.$$

Proof. Let

$$A_t := \left\{ |(1-\beta)\Xi(t) - X(t)| \le X(t)^{2/3} \right\}, \quad B_t := \bigcap_{k=t}^{\infty} A_k, \quad B := \bigcup_{t=1}^{\infty} B_t$$

$$C_t := \left\{ |X(t) - (1-\beta)\Xi(t)| \le \tfrac{1}{3}(1-\beta)^{2/3}\Xi(t)^{2/3} \right\}, \quad D_t := \bigcap_{k=t}^{\infty} C_k, \quad D := \bigcup_{t=1}^{\infty} D_t,$$

$$E_t := \left\{ (1-\beta)\Xi(t) > (t+1)^3(t+2)^3 \right\}, \quad J_t := \bigcap_{k=t}^{\infty} E_k, \quad J := \bigcup_{t=1}^{\infty} J_t.$$

By Theorems 1 and 2, we have $\mathbb{P}(J) = p_s$. For positive z, let $y_1(z)$ and $y_2(z)$ be the (unique) positive solution of the equations $z = y_1 + y_1^{2/3}$ and $z = y_2 - y_2^{2/3}$, respectively. Using y_1 and y_2, we write

$$A_t \cap E_t = \{ y_1((1-\beta)\Xi(t)) \le X(t) \le y_2((1-\beta)\Xi(t)) \} \cap E_t.$$

Note that if $z > 2$ we have $y_1 < z - z^{2/3}/3 < z + z^{2/3}/3 < y_2$. Hence, $C_t \cap E_t \subset A_t \cap E_t$ and, accordingly, $D_t \cap J_t \subset B_t \cap J_t$. By Chebyshev's inequality we have

$$\mathbb{P}(C_t | \Xi(t)) \ge 1 - 9\frac{\beta}{(1-\beta)^{1/3}}\Xi(t)^{-1/3},$$

which gives

$$\mathbb{P}(C_t^c | E_t) \le \frac{9\beta}{(t+1)(t+2)}. \tag{56}$$

For $t \ge \tau$, we have

$$\mathbb{P}(D_t^c \cap J_\tau) \le \sum_{k=t}^{\infty} \mathbb{P}(C_k^c \cap J_\tau) \le \sum_{k=t}^{\infty} \mathbb{P}(C_k^c \cap E_k) \le \sum_{k=t}^{\infty} \mathbb{P}(C_k^c | E_k) \le \sum_{k=t}^{\infty} \frac{9\beta}{(k+1)(k+2)} = \frac{9\beta}{t+1},$$

where used the definition of D_t for the first inequality, $J_\tau \subset E_k$ for the second inequality, $\mathbb{P}(E_k) \le 1$ for the third inequality, and (56) for the last inequality. Therefore, for any τ, we have $\mathbb{P}(D^c \cap J_\tau) = \lim_{t\to\infty} \mathbb{P}(D_t^c \cap J_\tau) = 0$ and, accordingly, $\mathbb{P}(D \cap J) = \mathbb{P}(B \cap J) = p_s$. As $x^{2/3} = x^{3/4}x^{-1/12}$ and $X(t)$ diverges almost surely on survival, the proof is completed. ☐

Remark 7.5. *Lemma 7.6 is applicable even if the support of* μ *is bounded because* $\Xi(t)$ *grows at least exponentially on survival. Therefore, regardless of the type of* G *we have almost surely on survival* $X(t) \sim (1-\beta)\Xi(t)$ *and the relative error of the approximation is at most* $\Xi(t)^{-1/4}$.

Definition (for Lemma 7.7). We define positive sequences $(a_t)_{t\geq 1}$ and $(y_t)_{t\geq 1}$ as

$$a_t := \frac{1}{\log(t+2)} - \frac{1}{\log(t+3)}, \quad y_t := \frac{\log a_t}{\log(1-\beta)}.$$

Note that a_t is monotonically decreasing with $1/a_t \sim t(\log t)^2$ and y_t is monotonically increasing. For $Y > y_t$, we define

$$\begin{aligned} \mathcal{W}_\mathrm{l}(Y,t) &:= \inf\left\{x > 0 : 1 - [1-\beta G(x)]^Y \leq a_t\right\}, \\ \mathcal{W}_\mathrm{s}(Y,t) &:= \sup\left\{x > 0 : [1-\beta G(x)]^Y \leq a_t\right\} - \varepsilon, \end{aligned}$$

where ε is an arbitrary small positive number. Since $a_t < \frac{1}{2}$, we have $\mathcal{W}_\mathrm{l}(Y,t) > \mathcal{W}_\mathrm{s}(Y,t)$. For $Y \leq y_t$ we define $\mathcal{W}_\mathrm{l}(Y,t) = \mathcal{W}_\mathrm{s}(Y,t) = 0$.

Remark 7.6. *Since $G(x)$ is a right-continuous-left-limit function, the purpose of introducing ε is to guarantee $(1-\beta G(\mathcal{W}_\mathrm{s}))^Y \leq a_t$. Without ε, $(1-\beta G(\mathcal{W}_\mathrm{s}))^Y$ may be larger than a_t. In the case that $G(x)$ is a strictly decreasing continuous function, we rather define, for $Y > y_t$,*

$$\begin{aligned} 1 - \beta G\left(\mathcal{W}_\mathrm{l}(Y,t)\right) &= (1-a_t)^{1/Y}, \\ 1 - \beta G\left(\mathcal{W}_\mathrm{s}(Y,t)\right) &= (a_t)^{1/Y}. \end{aligned}$$

Lemma 7.7. *For the type I tail function, define*

$$J := \left\{\lim_{t\to\infty} \frac{\log \Xi(t)}{t \log^{(n)}(t)} = \frac{1}{\alpha}\right\}.$$

For the type II tail function, define

$$J := \left\{\lim_{t\to\infty} \frac{\log^{(2)}(\Xi(t))}{\log t} = 1 + \frac{1}{\alpha}\right\},$$

for $n = 1$,

$$J := \left\{\lim_{t\to\infty} \frac{\log^{(3)}(\Xi(t))}{\log t} = \frac{1}{1+\alpha}\right\},$$

for $n = 2$, and

$$J := \left\{\lim_{t\to\infty} \frac{1}{\log^{(n-1)}(t)} \log\left(\frac{\log^{(2)}(\Xi(t))}{t}\right) = -\alpha\right\},$$

for $n \geq 3$. We have proved $\mathbb{P}(J) = p_s$. Let

$$A_t := \{\mathcal{W}_\mathrm{s}(\Xi(t),t) < W_t \leq \mathcal{W}_\mathrm{l}(\Xi(t),t)\}, \quad B_t := \bigcap_{k=t}^\infty A_k, \quad B := \bigcup_{t=1}^\infty B_t.$$

Then, $\mathbb{P}(B \cap J) = p_s$.

Proof. Let $C_t := \{\Xi(t) > y_t\}$, $D_\tau := \bigcap_{k=\tau}^\infty C_k$, and $D := \bigcup_{\tau=1}^\infty D_\tau$. Since $J \subset D$, it is enough to prove $\mathbb{P}(B \cap D) = p_s$, because $\mathbb{P}(B \cap J) = \mathbb{P}(B \cap D) - \mathbb{P}(B \cap (D \setminus J))$ and $\mathbb{P}(D \setminus J) = 0$. For any integer $Y > y_t$, we have

$$\begin{aligned}\mathbb{P}(A_t^c|\Xi(t) = Y) &= \mathbb{P}(W_t \le \mathcal{W}_\mathrm{s}|Y) + 1 - \mathbb{P}(W_t \le \mathcal{W}_\mathrm{l}|Y)\\ &= (1-\beta G(\mathcal{W}_\mathrm{s}))^Y + 1 - (1-\beta G(\mathcal{W}_\mathrm{l}))^Y \le 2a_t,\end{aligned}$$

where we have used Lemma 4.8. Accordingly, $\mathbb{P}(A_t^c|C_t) \le 2a_t$. If $t \ge \tau$, then we have

$$\mathbb{P}(B_t^c \cap D_\tau) \le \sum_{k=t}^\infty \mathbb{P}(A_k^c \cap D_\tau) \le \sum_{k=t}^\infty \mathbb{P}(A_k^c \cap C_k) \le \sum_{k=t}^\infty \mathbb{P}(A_k^c|C_k) \le \sum_{k=t}^\infty 2a_k = \frac{2}{\log(t+2)}.$$

Therefore, for any τ, we have $\mathbb{P}(B^c \cap D_\tau) = \lim_{t\to\infty} \mathbb{P}(B_t^c \cap D_\tau) = 0$ and so $\mathbb{P}(B^c \cap D) = 0$ and the proof is completed. □

Remark 7.7. *Assume G is a strictly decreasing continuous function. Let $w_0(Y,t)$ and $w_1(Y,t)$ be the solution of*

$$\begin{aligned}-\log G(w_1) &= \log Y + \log \tfrac{2\beta}{a_t},\\ -\log G(w_0) &= \log Y + \log\beta + \log^{(2)}\left(\tfrac{1}{a_t}\right).\end{aligned}$$

For sufficiently large t and Y, we have $1-(1-a_t)^{1/Y} > a_t/(2Y)$ and $1 - a_t^{1/Y} < -(\log a_t)/Y$, which gives $G(w_1) < G(\mathcal{W}_\mathrm{l}) < G(\mathcal{W}_\mathrm{s}) < G(w_0)$. Since G is a decreasing function, we have $w_0 \le \mathcal{W}_\mathrm{s} \le \mathcal{W}_\mathrm{l} \le w_1$. Hence, even if we define A_t by the condition $w_0 \le W_t \le w_1$ with $Y = \Xi(t)$, Lemma 7.7 remains valid.

Now consider $G(x) = \exp(-x^\alpha)$, which entails $\alpha W_t^\alpha \sim t\log t$. Then, we have

$$\begin{aligned}w_1 &= [\log\Xi(t)]^{1/\alpha}\left[1 + \frac{\log(2\beta/a_t)}{\log\Xi(t)}\right]^{1/\alpha},\\ w_0 &= [\log\Xi(t)]^{1/\alpha}\left[1 + \frac{\log\beta + \log^{(2)}(1/a_t)}{\log\Xi(t)}\right]^{1/\alpha}.\end{aligned}$$

We define a random sequence $(b_t)_{t\ge1}$ by $W_t = [\log X(t)]^{1/\alpha}\exp(b_t/t)$. Then, the above discussion together with Lemma 7.6 shows that, almost surely on survival,

$$0 \le \liminf_{t\to\infty} b_t \le \limsup_{t\to\infty} b_t \le 1.$$

Writing $W_t = u_1(t)e^{c_t}$ we see from Theorem 1 that, almost surely on survival, $\lim\limits_{t\to\infty} c_t = 0$. Hence,

$$\log X(t) = u_1(t)^\alpha \exp\left(\alpha c_t - \alpha\tfrac{b_t}{t}\right), \quad W_t = u_1(t)e^{c_t}. \tag{57}$$

Unfortunately, (57) *does not give an accurate estimate of $X(t)$, because*

$$X(t)\exp(-u_1(t)^\alpha) \ge \exp\left(\alpha u_1(t)^\alpha c_t - \alpha u_1(t)^\alpha \tfrac{b_t}{t}\right)$$

and it is unclear whether the right hand side will diverge or not. Since $X(t)$ rather than $\log X(t)$ is necessary to study the EFD and we do not have a tool to tame c_t and b_t, (57) *is too coarse to give a proof for Theorem 3 even for this special G. So we are forced to introduce simplified models, the DFMM and the SFMM, to prove variants of Theorem 3.*

7.2 Deterministic FMM and its EFD

Lemma 7.5 shows that if G is of type I with $n = 1$ or with $n = 2$ and $\alpha < 1$ then almost surely on survival

$$\lim_{t\to\infty} \frac{W_t}{u_n(t)} = \lim_{k\to\infty} \frac{N_k(k+s)}{(1-\beta)^s W_k^s} = 1,$$

for any nonnegative integer s. Hence, setting $W_k = u_n(k)$ and $N_k(t) = (1-\beta)^{t-k} u_n(k)^{t-k}$ for all large k and $t \geq k$ gives a good approximation of the models on survival. This approximation is especially convenient for the FMM. In this context, we are motivated to introduce the deterministic FMM as follows.

Definition of the DFMM. At each generation $k > 0$ a new mutant with fitness $W_k = u_n(k)$ appears. In case $u_n(k)$ is ill defined, we set $W_k = 1/(1-\beta)$. The number of non-mutated descendants of W_k grows deterministically as

$$N_k^{\mathrm{D}}(t) := (1-\beta)^{t-k} W_k^{t-k}. \tag{58}$$

where we neglect not only stochasticity but also the error due to the discreteness of $N_k(t)$. Note that in the DFMM only type I tail functions are under consideration and we make no restriction on n and α.

Notice that we added the superscript D in N_k^{D} to discern them from their stochastic counterparts. Since no fluctuation is present, the limit behavior of the EFD for the DFMM becomes a problem of calculus. In what follows, we will find a limit theorem of the EFD for the DFMM. To this end, we begin with the following elementary lemma.

Lemma 7.8. *Assume f is a positive continuous function that has a unique local maximum at x_c in a domain $[a-1, b+1]$, where a, b are integers and x_c need not be an integer. That is, $f(x) < f(y)$ if $a-1 \leq x < y \leq x_c$ and $f(x) > f(y)$ if $x_c \leq x < y \leq b+1$. Assume $a < x_c < b$. Let*

$$F(x) := \sum_{k=a}^{\lfloor x \rfloor} f(k).$$

Then, for any $a \leq x \leq b$,

$$\left| F(x) - \int_a^x f(y)dy \right| \leq 7 f(x_c).$$

Proof. Define $f_-(x) := f(\lfloor x \rfloor)$ and $f_+(x) = f(\lceil x \rceil)$. Then, for $a \leq x \leq b$,

$$F(x) = \int_a^{\lfloor x \rfloor + 1} f_-(y)dy = \int_{a-1}^{\lfloor x \rfloor} f_+(y)dy.$$

Note that $f_-(x) \leq f(x) \leq f_+(x)$ if $x < \lfloor x_c \rfloor$ and $f_-(x) \geq f(x) \geq f_+(x)$ if $x \geq \lceil x_c \rceil$. We abbreviate $m_c := \lfloor x_c \rfloor$ and $m := \lfloor x \rfloor$. If $m < m_c$, then

$$F(x) = \int_{a-1}^{m} f_+(y)dy \geq \int_{a-1}^{a} f(y)dy + \int_a^x f(y)dy - \int_m^x f(y)dy$$

and

$$F(x) = \int_a^{m+1} f_-(y)dy \leq \int_a^x f(y)dy + \int_x^{m+1} f(y),$$

which gives

$$\left| F(x) - \int_a^x f(y)dy \right| \leq 2 f(x_c).$$

For $m = m_c$, we use

$$\int_{a-1}^{a} f(y)dy + \int_{a}^{x} f(y)dy - \int_{m-1}^{x} f(y)dy \le F(m-1) \le \int_{a}^{x} f(y)dy - \int_{m}^{x} f(y)dy$$

and $F(x) = F(m-1) + f(m)$, to get $\left|F(x) - \int_a^x f(y)dy\right| \le 4f(x_c)$. If $m > m_c$, we consider

$$F(x) - F(x_c) = \sum_{k=m_c+1}^{m} f(k) = \int_{m_c+1}^{m+1} f_-(y)dy = f(m_c+1) + \int_{m_c+1}^{m} f_+(y)dy,$$

which gives

$$F(x) - F(x_c) \ge \int_{x_c}^{x} f(y)dy - \int_{x_c}^{m_c+1} f(y) + \int_{x}^{m+1} f(y)dy,$$
$$F(x) - F(x_c) \le f(x_c) + \int_{x_c}^{x} f(y)dy - \int_{x_c}^{m_c+1} f(y) - \int_{m}^{x} f(y)dy,$$

and, therefore, $\left|F(x) - F(x_c) - \int_{x_c}^x f(y)dy\right| \le 3f(x_c)$. Since

$$\left|F(x) - \int_{a}^{x} f(y)dy\right| \le \left|F(x_c) - \int_{a}^{x_c} f(y)dy\right| + \left|F(x) - F(x_c) - \int_{x_c}^{x} f(y)dy\right| \le 7f(x_c),$$

we have the desired result for any $a \le x \le b$. □

Definition. We define

$$H(x,t) := [(1-\beta)u_n(x)]^{t-x}, \quad h(x,t) := \log H(x,t),$$
$$\omega_1(x) := (1-\beta)\omega_W\left(\log^{(n-1)}(x)\right),$$
$$L_j(x) := (-1)^{j-1}\left(\frac{d}{dx}\right)^j \log^{(n)}(x), \quad \Omega_j(x) := (-1)^{j-1}\left(y\frac{d}{dy}\right)^j \log\omega_W(y)\bigg|_{y=\log^{(n-1)}(x)},$$

where $x \le t$ and x is assumed large enough so that the above definition makes sense. Note that $(1-\beta)u_n(x) = (\log^{(n-1)}(x))^{1/\alpha}\omega_1(x)$ and $N_k^{\mathrm{D}}(t) = H(k,t)$. Also note that

$$\frac{d}{dx}\log\omega_1(x) = L_1(x)\Omega_1(x), \quad \frac{d}{dx}\Omega_j(x) = -L_1(x)\Omega_{j+1}(x).$$

We assume $\lim_{x\to\infty}\Omega_j(x) = 0$ for any integer $j \ge 1$; see **(A4)**.

Lemma 7.9. *There are x_0 and t_0 such that*

$$\frac{\partial^2 h(x,t)}{\partial x^2} < 0, \quad \frac{\partial^3 h(x,t)}{\partial x^3} > 0, \quad \frac{\partial^4 h(x,t)}{\partial x^4} < 0, \tag{59}$$

for all $x \ge x_0 - 1$ and for all $t \ge t_0 \ge x_0 - 1$ with $x \le t$.

Proof. First observe that

$$L_1(x) = \left(\prod_{k=0}^{n-1}\log^{(k)}(x)\right)^{-1},$$
$$L_j(x) \sim \frac{(j-1)!L_1(x)}{x^{j-1}} = \frac{(j-1)!}{x^j}\left(\prod_{k=1}^{n-1}\log^{(k)}(x)\right)^{-1}, \quad \frac{L_j(x)}{L_{j+1}(x)} \sim \frac{x}{j}, \tag{60}$$

where we use the convention $\prod_{k=1}^{0} := 1$. We define

$$\phi_1(x,t) := 1 + \alpha\Omega_1(x) + \alpha\frac{L_1^2}{L_2}\Omega_2(x) + \frac{x}{t}\left[(1+\alpha\Omega_1(x))\left(\frac{2L_1}{L_2 x} - 1\right) - \alpha\frac{L_1^2}{L_2}\Omega_2(x)\right],$$
$$\phi_2(x,t) := \phi_1(x,t) - \frac{L_2}{L_3}\frac{\partial\phi_1}{\partial x}, \qquad \phi_3(x,t) := \phi_2(x,t) - \frac{L_3}{L_4}\frac{\partial\phi_2}{\partial x}.$$

We write down the derivatives

$$\begin{aligned}\frac{\partial h}{\partial x} &= -\frac{1}{\alpha}\log^{(n)}(x) - \log\omega_1(x) + (t-x)L_1(x)\left(\frac{1}{\alpha} + \Omega_1(x)\right), \frac{\partial^2 h}{\partial x^2} = -\frac{t}{\alpha}L_2(x)\phi_1(x,t),\\ \frac{\partial^3 h}{\partial x^3} &= \frac{t}{\alpha}L_3(x)\phi_2(x,t), \qquad \frac{\partial^4 h}{\partial x^4} = -\frac{t}{\alpha}L_4(x)\phi_3(x,t).\end{aligned} \tag{61}$$

As, by (60), $\phi_j(x,t)$ is positive for all sufficiently large x and t existence of x_0 and t_0 follows. □

Remark 7.8. *We fix such x_0 and t_0 in the following and treat x_0 as the initial generation and we consider only $t \geq t_0$.*

Lemma 7.10. *Let $x_c(t)$ be the location of the maximum of $h(x,t)$ for given t and let*

$$\kappa_t := -\left.\frac{\partial^2 h}{\partial x^2}\right|_{x=x_c}, \quad d_t := \frac{1}{3!}\left.\frac{\partial^3 h}{\partial x^3}\right|_{x=x_c}.$$

Then,

$$x_c \sim t\prod_{k=1}^{n}\frac{1}{\log^{(k)}(t)}, \tag{62}$$
$$\kappa_t \sim \frac{\log^{(n)}(t)}{\alpha t}\prod_{k=1}^{n}\log^{(k)}(t) \sim \frac{\log^{(n)}(t)}{\alpha x_c}, \quad d_t \sim \frac{\log^{(n)}(t)}{3\alpha t^2}\prod_{k=1}^{n}\left(\log^{(k)}(t)\right)^2. \tag{63}$$

Proof. From (61) and (59), we have

$$0 = -\frac{1}{\alpha}\log^{(n)}(x_c) - \log\omega_1(x_c) + (t-x_c)L_1(x_c)\left(\frac{1}{\alpha} + \Omega_1(x_c)\right),$$

for given t. Obviously, the solution of the equation diverges with t, so x_c satisfies

$$t \sim \frac{\log^{(n)}(x_c)}{L_1(x_c)} = x_c\prod_{k=1}^{n}\log^{(k)}(x_c).$$

Therefore,

$$x_c \sim t\prod_{k=1}^{n}\frac{1}{\log^{(k)}(x_c)} \sim t\prod_{k=1}^{n}\frac{1}{\log^{(k)}(t)}.$$

Considering $\phi_j(x_c) \sim 1$ and using (60), we get the desired result. □

Remark 7.9. *In the following, t_0 is further assumed so large that $x_c > x_0$ for all $t > t_0$.*

Lemma 7.11.

$$|h(x,t) - h(x_c,t)| \leq \frac{\kappa_t}{2}(x-x_c)^2\left(1 + \frac{2d_t}{\kappa_t}|x-x_c|\right).$$

Proof. By (59), we have

$$-\frac{\kappa_t}{2}(x-x_c)^2 + d_t(x-x_c)^3 \le h(x,t) - h(x_c,t) \le -\frac{\kappa_t}{2}(x-x_c)^2$$

for $x_0 \le x \le x_c$ and

$$-\frac{\kappa_t}{2}(x-x_c)^2 + d_t(x-x_c)^3 \ge h(x,t) - h(x_c,t) \ge -\frac{\kappa_t}{2}(x-x_c)^2 \tag{64}$$

for $x \ge x_c$, and, therefore, we get the desired result. □

Lemma 7.12. *We define*

$$\widetilde{X}(t) := \sum_{k=x_0}^{t} N_k^{\mathrm{D}}(t), \quad \Phi(f,t) := \frac{1}{\widetilde{X}(t)} \sum_{k=x_0}^{t} N_k^{\mathrm{D}}(t)\Theta(f - u_n(k)),$$
$$S_t^{\mathrm{D}} := \frac{1}{\widetilde{X}(t)} \sum_{k=x_0}^{t} u_n(k) N_k^{\mathrm{D}}(t), \quad \sigma_t^{\mathrm{D}} := \left(\frac{1}{\widetilde{X}(t)} \sum_{k=x_0}^{t} (u_n(k) - S_t^{\mathrm{D}})^2 N_k^{\mathrm{D}}(t)\right)^{1/2}, \tag{65}$$

where we only consider $t > t_0$. Then,

$$S_t^{\mathrm{D}} \sim v_n(t), \quad \sigma_t^{\mathrm{D}} \sim \mathfrak{s}_n(t), \quad \lim_{t\to\infty} \Phi(v_n(t) + \mathfrak{s}_n(t) y, t) = \lim_{t\to\infty} \Phi(S_t^{\mathrm{D}} + \sigma_t^{\mathrm{D}} y, t) = \Upsilon(y),$$

where

$$v_n(t) := \alpha^{-\delta_{n,1}/\alpha} \left(\log^{(n-1)}(t)\right)^{1/\alpha} L\left(\left(\log^{(n-1)}(t)\right)^{1/\alpha}\right), \tag{66}$$

$$\mathfrak{s}_n(t) := \frac{v_n(t)}{\sqrt{\alpha t}} \left(\prod_{k=1}^{n-1} \log^{(k)}(t)\right)^{-1/2}. \tag{67}$$

Proof. First note that (63) gives, for any $0 < \varepsilon < 1$,

$$\lim_{t\to\infty} \frac{2d_t}{\kappa_t} \kappa_t^{-(1-\varepsilon)/2} = 0,$$

If $|x - x_c| \le \kappa_t^{-(1-\varepsilon)/2}$ in Lemma 7.11 for some $0 < \varepsilon < 1$ and t is sufficiently large that $2d_t\kappa_t^{-(1-\varepsilon)/2}/\kappa_t \le 1$, then $|h(x,t) - h(x_c,t)| \le \kappa_t^{\varepsilon}$, which approaches zero as t goes to infinity; see (63). Therefore, $h(m,t) \sim h(x_c,t)$, as $t \to \infty$ for $|m - x_c| \le \kappa_t^{-(1-\varepsilon)/2}$, which gives

$$\lim_{t\to\infty} \kappa_t^{(1-\varepsilon)/2} \frac{\widetilde{X}(t)}{H(x_c,t)} = \infty.$$

for any $\varepsilon > 0$. In other words, for any $\varepsilon > 0$ there is t_1 such that $\widetilde{X}(t) \ge H(x_c,t)\kappa_t^{-(1-\varepsilon)/2}$ for all $t \ge t_1$, which, along with Lemma 7.8, gives

$$\lim_{t\to\infty} \left| \Phi(u_n(z),t) - \frac{1}{\widetilde{X}(t)} \int_{x_0}^{z} H(y,t) dy \right| = 0, \tag{68}$$

where z should be regarded as a certain monotonically increasing function of t with $x_0 < z \le t$.

Now consider the other case. Fix $0 < \varepsilon < 1$ and define $x_\pm := x_c \pm \kappa_t^{-(1+\varepsilon)/2}$ and also $z_\pm := x_c \pm 2\kappa_t^{-(1+\varepsilon)/2}$. By (59), we always have $h(x,t) \le h(x_\pm, t) + \xi_\pm(x - x_\pm)$, where

$$\xi_\pm = \left.\frac{\partial h}{\partial x}\right|_{x=x_\pm} = -\frac{1}{\alpha}\log^{(n)}(x_\pm) - \log\omega_1(x_\pm) + (t - x_\pm)L_1(x_\pm)\left(\frac{1}{\alpha} + \Omega_1(x_\pm)\right).$$

Since $\lim_{t\to\infty}\kappa_t^{-(1+\varepsilon)/2}/x_c = 0$, Taylor's theorem gives

$$\xi_\pm \sim \pm\left(\left.\frac{\partial^2 h}{\partial x^2}\right|_{x=x_c}\right)\kappa_t^{-(1+\varepsilon)/2} = \mp\kappa_t^{(1-\varepsilon)/2}, \quad \xi_\pm(z_\pm - x_\pm) \sim -\kappa_t^{-\varepsilon}.$$

Now consider

$$\begin{aligned}
I_1(t) &:= \int_{x_0}^{z_-}\frac{H(y,t)}{H(x_c,t)}dy \le \int_{x_0}^{z_-} e^{h(y,t)-h(x_-,t)}dy \\
&\le \int_{-\infty}^{z_-} e^{\xi_-(y-x_-)}dy \sim \kappa_t^{-(1-\varepsilon)/2}\exp\left(-\kappa_t^{-\varepsilon}\right), \\
I_2(t) &:= \int_{z_+}^{t}\frac{H(y,t)}{H(x_c,t)}dy \le \int_{z_+}^{t} e^{h(y,t)-h(x_+,t)}dy \\
&\le \int_{z_+}^{\infty} e^{\xi_+(y-x_+)}dy \sim \kappa_t^{-(1-\varepsilon)/2}\exp\left(-\kappa_t^{-\varepsilon}\right),
\end{aligned} \tag{69}$$

where we have used $H(x_\pm,t) \le H(x_c,t)$. Since $\widetilde{X}(t) \ge H(x_c,t)$ for all sufficiently large t and $\lim_{t\to\infty} I_1(t) = \lim_{t\to\infty} I_2(t) = 0$, (68) yields, for any $\varepsilon > 0$,

$$\lim_{t\to\infty}\Phi(u_n(z),t) = \begin{cases} 0, & z \le x_c - \kappa_t^{-(1+\varepsilon)/2}, \\ 1, & z \ge x_c + \kappa_t^{-(1+\varepsilon)/2}. \end{cases}$$

Hence, it is enough to consider $\Phi(u_n(z),t)$ for $|x_c - z| \le \kappa_t^{-(1+\varepsilon)/2}$ for a certain positive ε.

Abbreviate $z := x_c + y/\sqrt{\kappa_t}$ and assume $|y| \le \kappa_t^{-1/8}$ (in a sense, we have set $\varepsilon = 1/4$). By Taylor's theorem, there is y_0 such that $|y_0| \le |y|$ and

$$h(z,t) = h(x_c,t) - \frac{1}{2}y^2 + R_t\left(x_c + \frac{y_0}{\sqrt{\kappa_t}}\right)y^3, \quad R_t(x) := \frac{t}{6\alpha}L_3(x)\phi_2(x,t).$$

Defining

$$\varepsilon_1(t) = \exp\left(\sup\left\{\left|R_t\left(x_c + \frac{y_0}{\sqrt{\kappa_t}}\right)y^3\right| : |y| \le \kappa_t^{-1/8}\right\}\right) - 1,$$

we have

$$\frac{H(z,t)}{H(x_c,t)} \simeq_{\varepsilon_1(t)} \exp\left(-\frac{y^2}{2}\right), \tag{70}$$

where $A \simeq_\varepsilon B$ is a shorthand notation for $(1-\varepsilon)B \le A \le (1+\varepsilon)B$. Then,

$$\int_{x_c - a_t^{-5/8}}^{z}\frac{H(x,t)}{H(x_c,t)}dx \simeq_{\varepsilon_1(t)} \kappa_t^{-1/2}\int_{-\kappa_t^{-1/8}}^{y}\exp\left(-\frac{x^2}{2}\right)dx, \tag{71}$$

where $\kappa_t^{-1/2}$ is the Jacobian of the change of variables. Since $R_t \sim d_t \sim \kappa_t/(3x_c)$ and, accordingly, $\lim_{t\to\infty} \varepsilon_1(t) = 0$, we have

$$\lim_{t\to\infty} \frac{\widetilde{X}(t)\sqrt{\kappa_t}}{H(x_c,t)} = \lim_{t\to\infty} \int_{-\kappa_t^{-1/8}}^{\kappa_t^{-1/8}} e^{-x^2/2}dx = \int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}, \tag{72}$$

which, together with (68), gives

$$\lim_{t\to\infty} \Phi(u_n(x_c + y/\sqrt{\kappa_t}), t) = \Upsilon(y). \tag{73}$$

To complete the proof, we have to show

$$\lim_{t\to\infty} \Phi\left(u_n\left(x_c + \frac{y}{\sqrt{\kappa_t}}\right), t\right) = \lim_{t\to\infty} \Phi(S_t^{\mathrm{D}} + \sigma_t^{\mathrm{D}} y, t),$$

for $|y| \le \kappa_t^{-1/8}$. Let $S_t' := u_n(x_c)$ and let y_c be a function of t implicitly defined as the solution of the equation

$$\left.\frac{\partial h_2(x,t)}{\partial x}\right|_{x=y_c} = L_1(y_c)\left(\nu + \Omega_2^{(1)}(y_c)\right) + \left.\frac{\partial h(x,t)}{\partial x}\right|_{x=y_c},$$

where $h_2(x,t) := \log u_n(x) + h(x,t) = \log(u_n(x)H(x,t))$. Notice that $u_n(x_c) \sim v_n(t)$. Obviously, $y_c \sim x_c$. Define

$$\rho_1(t) := \frac{S_t^{\mathrm{D}}}{S_t'} = \frac{1}{\widetilde{X}(t)S_t'} \sum_{k=x_0}^{t} u_n(k) N_k^{\mathrm{D}}(t) = \frac{1}{\widetilde{X}(t)S_t'} \sum_{k=x_0}^{t} e^{h_2(k,t)}.$$

Since $h_2(x,t)$ for given t satisfies the condition in Lemma 7.8, we have

$$\left|\rho_1(t) - \frac{1}{\widetilde{X}(t)S_t'} \int_{x_0}^{t} u_n(y)H(y,t)dy\right| \le \frac{7H(y_c,t)u_n(y_c)}{\widetilde{X}(t)S_t'}.$$

Since $\lim_{t\to\infty} H(y_c,t)/\widetilde{X}(t) = 0$ and $u_n(y_c)/S_t' \sim 1$, we have

$$\lim_{t\to\infty} \left|\rho_1(t) - \frac{1}{\widetilde{X}(t)S_t'} \int_{x_0}^{t} u_n(y)H(y,t)dy\right| = 0.$$

Let $z_\pm = x_c \pm \kappa_t^{-5/8}$. Since

$$\int_{z_+}^{t} u_n(y)^m \frac{H(y,t)}{H(x_c,t)} dy \le u_n(t)^m \int_{z_+}^{t} \frac{H(y,t)}{H(x_c,t)} dy,$$
$$\int_{x_0}^{z_-} u_n(y)^m \frac{H(y,t)}{H(x_c,t)} dy \le u_n(t)^m \int_{x_0}^{z_-} \frac{H(y,t)}{H(x_c,t)} dy,$$

I_1 and I_2 in (69) with $\varepsilon = 1/4$ yield

$$\lim_{t\to\infty} \frac{1}{\widetilde{X}(t)S_t'} \left|\int_{x_0}^{t} u_n(y)^m H(y,t)dy - \int_{z_-}^{z_+} u_n(y)^m H(y,t)dy\right| = 0,$$

for any positive integer m. Using (70), we have

$$\frac{1}{\widetilde{X}(t)S'_t}\int_{z_-}^{z_+} u_n(z)H(z,t)dz \simeq_{\varepsilon_1(t)} \frac{H(x_c,t)\kappa_t^{-1/2}}{\widetilde{X}(t)S'_t}\int_{-\kappa_t^{-1/8}}^{\kappa_t^{-1/8}} u_n\left(x_c+\frac{y}{\sqrt{\kappa_t}}\right)e^{-y^2/2}dy.$$

Since $S'_t \sim u_n(x_c + y/\sqrt{\kappa_t})$, we have

$$\lim_{t\to\infty}\frac{1}{\widetilde{X}(t)S'_t}\int_{z_-}^{z_+} u_n(y)H(y,t)dy = 1.$$

Therefore $\rho_1(t) \sim 1$ or $S_t^{\mathrm{D}} \sim u_n(x_c) \sim v_n(t)$, as claimed.

Define

$$\begin{aligned}
\sigma'_t &:= \kappa_t^{-1/2}\left.\frac{du_n}{dx}\right|_{x=x_c} = \frac{S'_t}{\sqrt{\kappa_t}}L_1(x_c)\left[\frac{1}{\alpha}+\Omega_1(x_c)\right],\\
\rho_2(t) &:= \frac{S_t^{\mathrm{D}}-S'_t}{\sigma'_t} = \frac{1}{\widetilde{X}(t)}\sum_{k=x_0}^{t}\frac{u_n(k)-u_n(x_c)}{\sigma'_t}H(k,t),\\
\rho_3(t) &:= \frac{1}{\widetilde{X}(t)}\int_{z_-}^{z_+}\frac{u_n(x)-u_n(x_c)}{\sigma'_t}H(x,t)dx.
\end{aligned}$$

Note that $\sigma'_t \sim \mathfrak{s}_n(t)$. Assume $|y| \le \kappa_t^{-1/8}$. By Taylor's theorem, there is y_1 with $|y_1| \le |y|$ such that

$$\frac{u_n(x_c+y/\sqrt{\kappa_t})-u_n(x_c)}{\sigma'_t} = y + \frac{\tilde{R}_t(x_c+y_1/\sqrt{\kappa_t})}{\sigma'_t}y^2,$$

where

$$\tilde{R}_t(x) := \frac{1}{2\kappa_t}\frac{d^2u_n(x)}{dx^2} = \frac{u_n(x)}{2\kappa_t}L_2(x)\left(\frac{L_1^2}{L_2}\left[\left(\frac{1}{\alpha}+\Omega_1(x)\right)^2-\Omega_2(x)\right]-\frac{1}{\alpha}-\Omega_1(x)\right).$$

Using

$$\frac{\tilde{R}_t(x_c+y_1/\sqrt{\kappa_t})}{\sigma'_t} \sim \frac{\tilde{R}_t(x_c)}{\sigma'_t} \sim \frac{1}{2\sqrt{\kappa_t x_c^2}}\left(\frac{\delta_{n,1}}{\alpha}-1\right) \tag{74}$$

for $|y_1| \le \kappa_t^{-1/8}$, $\int_{-x}^{x} ye^{-y^2/2}dy = 0$, and (70), we have

$$|\rho_3(t)| \simeq_{\varepsilon(t)} \frac{\kappa_t^{-1/2}H(x_c,t)}{\widetilde{X}(t)}\frac{1}{2\sqrt{\kappa_t x_c^2}}\left|\frac{\delta_{n,1}}{\alpha}-1\right|\int_{-\kappa_t^{-1/8}}^{\kappa_t^{-1/8}} y^2e^{-y^2/2}dy,$$

where $\lim_{t\to\infty}\varepsilon(t) = 0$. Therefore,

$$\lim_{t\to\infty}\rho_3(t) = 0. \tag{75}$$

Since

$$\begin{aligned}
\left|\frac{1}{\widetilde{X}(t)}\int_{x_0}^{z_-}\frac{u_n(x)-u_n(x_c)}{\sigma'_t}H(x,t)dx\right| &\le \frac{2u_n(t)}{\sigma'_t}\left|\frac{1}{\widetilde{X}(t)}\int_{x_0}^{z_-}H(x,t)dx\right|,\\
\left|\frac{1}{\widetilde{X}(t)}\int_{z_+}^{t}\frac{u_n(x)-u_n(x_c)}{\sigma'_t}H(x,t)dx\right| &\le \frac{2u_n(t)}{\sigma'_t}\left|\frac{1}{\widetilde{X}(t)}\int_{z_+}^{t}H(x,t)dx\right|,
\end{aligned}$$

(69) together with (75) gives

$$\lim_{t\to\infty}\rho_2(t)=\lim_{t\to\infty}\frac{1}{\widetilde{X}(t)}\int_{z_-}^{z_+}\frac{u_n(x)-u_n(x_c)}{\sigma_t'}H(x,t)dx=0. \tag{76}$$

Define

$$\begin{aligned}
\rho_4(t):=&\frac{\sigma_t^{\mathrm{D}\,2}}{(\sigma_t')^2}=\frac{1}{\widetilde{X}(t)(\sigma_t')^2}\sum_k\left(u_n(k)-S_t'-\sigma_t'\rho_2(t)\right)^2H(k,t)\\
=&\frac{1}{\widetilde{X}(t)}\sum_k\left(\frac{u_n(k)-u_n(x_c)}{\sigma_t'}\right)^2H(k,t)-\rho_2(t)^2,\\
\rho_5(t):=&\frac{1}{\widetilde{X}(t)}\int_{z_-}^{z_+}\left(\frac{u_n(x)-u_n(x_c)}{\sigma_t'}\right)^2H(x,t)dx\\
=&\frac{1}{\widetilde{X}(t)}\int_{z_-}^{z_+}\kappa_t(x-x_c)^2\left(1+\frac{\tilde{R}_t(x_c+y_1/\sqrt{\kappa_t})}{\sigma_t'}\right)^2H(x,t)dx.
\end{aligned}$$

Using (70), (72), (74), and (76), we have

$$\lim_{t\to\infty}\rho_4(t)=\lim_{t\to\infty}\rho_5(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y^2e^{-y^2/2}dy=1,$$

where we have also used the same procedure to arrive at (76) using $(u_n(x)-u_n(x_c))^2\le 4u_n(t)^2$. From the above calculations, we conclude that there is a constant C such that

$$|\rho_2(t)|\le\frac{C}{\sqrt{\kappa_t}x_c},\quad |\rho_4(t)-1|\le\frac{C}{\sqrt{\kappa_t}x_c}, \tag{77}$$

for all sufficiently large t.

Let $z:=x_c+y/\sqrt{\kappa_t}$ and $z':=u_n^{-1}(S_t^{\mathrm{D}}+\sigma_t^{\mathrm{D}}y)$. Recall that for any small but positive ε_2 and ε_3, $\widetilde{X}(t)\ge\kappa_t^{-(1-\varepsilon_2)/2}H(x_c,t)$ and $\kappa_t\le t^{-1+\varepsilon_3}$ for all sufficiently large t. Since

$$\lim_{t\to\infty}|\Phi(u_n(z),t)-\Phi(u_n(z'),t)|=\lim_{t\to\infty}\frac{1}{\widetilde{X}(t)}\left|\int_z^{z'}H(x,t)dx\right|\le\lim_{t\to\infty}t^{-(1-\varepsilon_0)/2}|z-z'|,$$

for any $0<\varepsilon_0<1$, we need to show that there is ε_0 such that $\lim_{t\to\infty}t^{-(1-\varepsilon_0)/2}|z-z'|=0$. First observe that $S_t^{\mathrm{D}}+\sigma_t^{\mathrm{D}}y=S_t'+\sigma_t'y'$ for $y':=\rho_2(t)+y\sqrt{\rho_4(t)}$. Assume t is so large that $|y'|\le 2\kappa_t^{-1/8}$. By Taylor's theorem, there is y_1 such that $|y_1|\le|y'|\le 2\kappa_t^{-1/8}$ and

$$\begin{aligned}
z'&=u_n^{-1}(S_t')+\frac{\sigma_t'}{u_n'(z_1)}y'\\
&=z+\left(\frac{\sigma_t'}{u_n'(z_1)}-\frac{1}{\sqrt{\kappa_t}}\right)y+\frac{\sigma_t'}{u_n'(z_1)}\left[\rho_2(t)+y(\sqrt{\rho_4}-1)\right],
\end{aligned}$$

where $z_1=u_n^{-1}(S_t'+\sigma_t'y_1)$. Using $z_1\sim x_c$, $u_n(x_c)=\sigma_t'\sqrt{\kappa_t}$, (77), and $\lim_{t\to\infty}t^{-\varepsilon_4}/(\kappa_tx_c)=0$ for any $\varepsilon_4>0$, we have $|z'-z|\le\kappa_t^{-1/6}\le t^{1/4}$ for all sufficiently large t. Hence, if we choose $\varepsilon_0=1/8$, we have the desired result. Since $\rho_4(t)\sim 1$, the proof is completed. □

Remark 7.10. *Since* $u_n(t)/v_n(t) \sim (\log t)^{\delta_{n,1}/\alpha}$, *we have*

$$S_t^{\mathrm{D}} \sim W_t \ n \geq 2, \text{ while}$$
$$\lim_{t\to\infty} S_t^{\mathrm{D}}/W_t = 0 \text{ for } n = 1.$$

In other words, when $n \geq 2$, *the mean fitness at generation* t *is hardly discernible from the largest fitness at the same generation. Another interesting observation is that if* $n \geq 2$ *or if* $n = 1$ *and* $\alpha > 2$, *then* $\lim_{t\to\infty} \sigma_t^{\mathrm{D}} = 0$, *which implies that the width of the traveling wave decreases to zero and the EFD becomes a delta function in the sense that*

$$\lim_{t\to\infty} \Phi(S_t^{\mathrm{D}} + y, t) = \begin{cases} 0, & y < 0, \\ 1, & y > 0. \end{cases}$$

This should be compared with the case of $n = 1$ *and* $\alpha < 2$ *in which the width of the traveling wave increases with generation. For* $n = 1$ *and* $\alpha = 2$, *the behaviour of* σ_t^{D} *depends on the slowly varying function* L *entering the tail function in* (1).

7.3 Semi-deterministic FMM and its EFD

Definition of the SFMM. At each generation $k \geq 0$ a new mutant with fitness

$$\theta_k := (1-\beta)u_n(k)$$

appears and $(N_k(t)\colon t \geq k)$ are mutually independent Galton-Watson processes with Poisson-distributed offspring with mean θ_k for each k. In case $u_n(k)$ is ill-defined, we set $\theta_k = 1$. By definition, $N_k(k) = 1$ and $N_k(\tau) = 0$ for $\tau < k$ and no extinction is possible in the SFMM. Since we will use Lemma 7.5 to prove Theorem 4 below, we limit the definition of the SFMM to the case $n = 1$ or the case $n = 2$ and $\alpha < 1$; see also Remark 7.2.

We denote the total population size of the SFMM at generation t by

$$X^{\mathrm{S}}(t) := \sum_{k=0}^{t} N_k(t).$$

The EFD $\Psi_s(f,t)$ of the SFMM and its mean fitness S_t are defined as

$$\Psi_s(f,t) := \frac{1}{X^{\mathrm{S}}(t)} \sum_{k=0}^{t} N_k(t)\Theta(f - u_n(k)), \quad S_t^{\mathrm{S}} := \frac{1}{X^{\mathrm{S}}(t)} \sum_{k=0}^{t} u_n(k) N_k(t).$$

Since $N_k(t)$ is the number of non-mutated descendants, we put $(1-\beta)$ in the definition of the fitness of a new mutant in the SFMM. In a sense, the SFMM is closer to the FMM than the DFMM due to fluctuations of $N_k(t)$. We redefine $u_n(k) := \theta_k/(1-\beta)$ for convenience. Now we prove that the EFD of the SFMM in the long time limit becomes almost surely a Gaussian traveling wave just as the DFMM.

Theorem 4. *For the SFMM with* $n = 1$ *or with* $n = 2$ *and* $\alpha < 1$, *almost surely*

$$\lim_{t\to\infty} \Psi_s(v_n(t) + \mathfrak{s}_n(t)y, t) = \Upsilon(y), \quad \lim_{t\to\infty} \frac{S_t^{\mathrm{S}}}{v_n(t)} = 1,$$

where

$$v_n(t) = \alpha^{-\delta_{n,1}/\alpha} \left(\log^{(n-1)}(t)\right)^{1/\alpha} L\left(\left(\log^{(n-1)}(t)\right)^{1/\alpha}\right),$$

$$\mathfrak{s}_n(t) = \frac{v_n(t)}{\sqrt{\alpha t}} \left(\prod_{k=1}^{n-1} \log^{(k)}(t)\right)^{-1/2}$$

have been introduced previously in (66) *and* (67).

Proof. We define J and E as in Lemma 7.5. It is obvious that Lemma 7.5 is applicable to the SFMM. Note that by definition J in (52) for the SFMM can be regarded as the sample space and, accordingly, $\mathbb{P}(E) = 1$. For any $0 < \varepsilon < 1/2$ and for any outcome $\omega \in E$, there exists τ_1 such that $(1-\varepsilon)\theta_k^{t-k} \le N_k(t) \le (1+\varepsilon)\theta_k^{t-k}$ for all $t \ge k \ge \tau_1$. Notice that τ_1 can vary from outcome to outcome. Let

$$X^{\mathrm{S}}(t, \tau_1) := \sum_{k=0}^{\tau_1} N_k(t), \quad X^{\mathrm{D}}(t) := \sum_{k=0}^{t} \theta_k^{t-k}, \quad X^{\mathrm{D}}(t, \tau_1) := \sum_{k=0}^{\tau_1} \theta_k^{t-k}.$$

Then, for $t \ge \tau_1$, we have

$$(1-\varepsilon)\left(X^{\mathrm{D}}(t) - X^{\mathrm{D}}(t,\tau_1)\right) + X^{\mathrm{S}}(t,\tau_1) \le X^{\mathrm{S}}(t) \le (1+\varepsilon)\left(X^{\mathrm{D}}(t) - X^{\mathrm{D}}(t,\tau_1)\right) + X^{\mathrm{S}}(t,\tau_1).$$

Since $X_s^{\mathrm{D}}(t,\tau_1)$ and $X^{\mathrm{D}}(t,\tau_1)$ grow at most exponentially and $X^{\mathrm{D}}(t)$ grows super-exponentially, we have almost surely

$$\liminf_{t\to\infty} \frac{X^{\mathrm{S}}(t)}{X^{\mathrm{D}}(t)} \ge 1-\varepsilon, \quad \limsup_{t\to\infty} \frac{X^{\mathrm{S}}(t)}{X^{\mathrm{D}}(t)} \le 1+\varepsilon.$$

Hence there is almost surely τ_2 such that $(1-2\varepsilon)X^{\mathrm{D}}(t) \le X^{\mathrm{S}}(t) \le (1+2\varepsilon)X^{\mathrm{D}}(t)$ for all $t \ge \tau_2$.

Now set $\tau = \max\{\tau_1, \tau_2\}$ and assume $t > \tau$. Then, we have

$$\begin{aligned}\Psi_s(f,t) &\ge \frac{1}{1+2\varepsilon}\frac{1}{X^{\mathrm{D}}(t)}\sum_{k=0}^{\tau_1} N_k(t)\Theta(f-u_n(k)) + \frac{1}{1+2\varepsilon}\frac{1}{X^{\mathrm{D}}(t)}\sum_{k=\tau_1+1}^{t} N_k(t)\Theta(f-u_n(k)) \\ &\ge \frac{1}{1+2\varepsilon}\frac{1}{X^{\mathrm{D}}(t)}\sum_{k=0}^{\tau_1} N_k(t)\Theta(f-u_n(k)) + \frac{1-\varepsilon}{1+2\varepsilon}\frac{1}{X^{\mathrm{D}}(t)}\sum_{k=\tau_1+1}^{t} \theta_k^{t-k}\Theta(f-u_n(k)).\end{aligned}$$

Hence by Lemma 7.12, we conclude

$$\liminf_{t\to\infty} \Psi_s(v_n(t) + \mathfrak{s}_n(t)y, t) \ge \frac{1-\varepsilon}{1+2\varepsilon}\Upsilon(y).$$

By the same token, we have

$$\limsup_{t\to\infty} \Psi_s(v_n(t) + \mathfrak{s}_n(t)y, t) \le \frac{1+\varepsilon}{1-2\varepsilon}\Upsilon(y).$$

Since ε is arbitrary, we proved the first part of the theorem.

Let

$$S_t^{\mathrm{D}} := \frac{1}{X^{\mathrm{D}}(t)} \sum_{k=0}^{t} u_n(k)\theta_k^{t-k}.$$

By Lemma 7.12, we have $S_t^{\mathrm{D}} \sim v_n(t)$. Inspecting the above proof, we can conclude that for any $0 < \varepsilon < 1/2$ and for any outcome $\omega \in E$, there is τ such that

$$(1-\varepsilon)S_t^{\mathrm{D}} \le S_t^{\mathrm{S}} \le (1+\varepsilon)S_t^{\mathrm{D}},$$

for all $t \ge \tau$. Since ε is arbitrary, the proof is completed. □

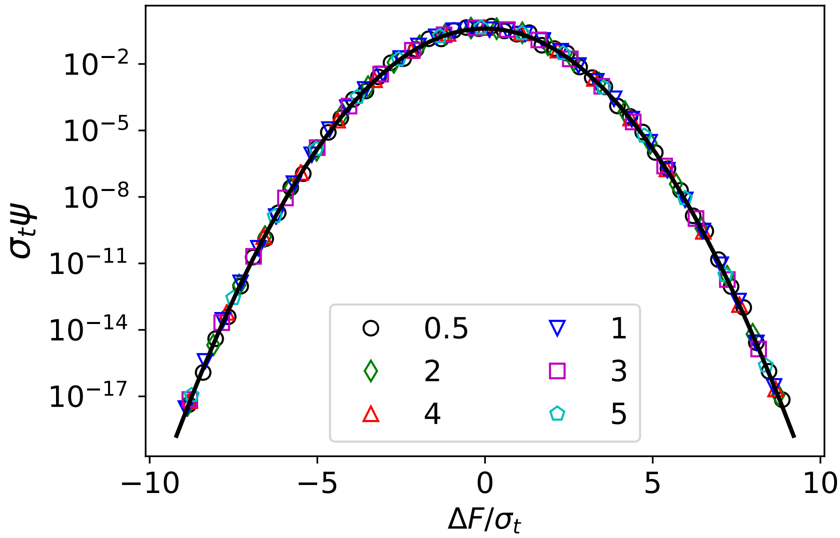

Figure 1: Semilogarithmic plot of $\sigma_t\psi(F,t)$ vs. $\Delta F/\sigma_t$ for various α's at $t = 983\ 040$. For comparison, the normal distribution is plotted by a solid curve.

7.4 Numerical study for the MMM with $n = 1$

Since the largest fitness is expected to dominate the evolution of the population even in the MMM and the limiting distribution is continuous even in the FMM, we conjecture that Theorem 3 is valid even for the MMM; see Remark 2.1. For the MMM, however, we only present some numerical results, which supports our conjecture.

For numerical feasibility we assume that the fitness of a mutant can only be one of the discrete values $f_i = (ci)^{1/\alpha}$ for $i \geq 1$, where c is a constant to be determined later. Defining $G_p(x) = \exp(-x^\alpha + c)$ for $x \geq c^{1/\alpha}$ and $G_p(x) = 1$ for $x \leq c^{1/\alpha}$, we assign probabilities

$$p_i := \mathbb{P}(F = f_i) = G_p(f_i) - G_p(f_{i+1}) = e^{-ci}(e^c - 1).$$

Since $G_p(f_{i+1}) \leq G(x) \leq G_p(f_i)$ for $f_i \leq x < f_{i+1}$ and $\lim_{i\to\infty} f_{i+1}/f_i = 1$, we have

$$\lim_{x\to\infty} \frac{\log G(x)}{x^\alpha} = -1.$$

Therefore, we can apply Theorem 1, to predict $W_t \sim \alpha^{-1/\alpha}(t\log t)^{1/\alpha}$, almost surely on survival.

In this section, we denote the number of individuals with fitness f_k at generation t by $N_k(t)$. We would like to emphasize that f_k should not be confused with W_k. The total population size $X(t)$ and the mean fitness S_t are calculated as

$$X(t) = \sum_{k=1}^{\infty} N_k(t), \quad S_t = \sum_{k=1}^{\infty} \frac{N_k(t)}{X(t)} f_k. \tag{78}$$

The standard deviation σ_t is naturally defined. Given $N_k(t)$ and S_t, the random variable $N_k(t+1)$ is drawn from the Poisson distribution with mean $(1-\beta)N_k(t)f_k + \beta S_t X(t) p_k$. Since the accurate value of β is not important as long as $0 < \beta < 1$, we choose $\beta = 10^{-20}$ to make $1-\beta$ indistinguishable from 1 within machine accuracy of double-precision floating-point format.

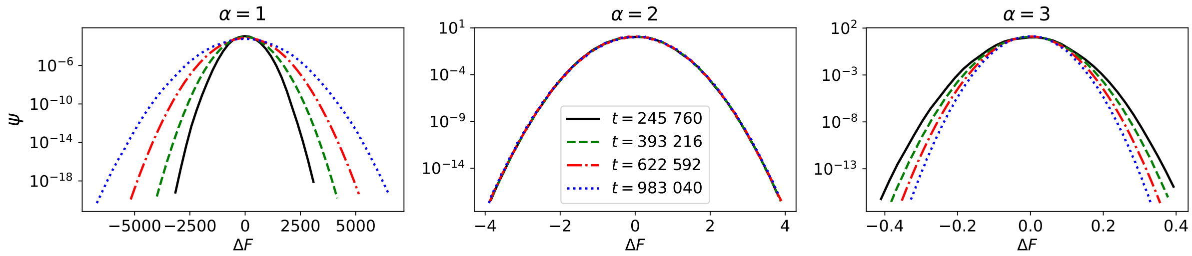

Figure 2: Plots of $\psi(F,t)$ vs. ΔF at different generations for $\alpha = 1$ (left), $\alpha = 2$ (middle), and $\alpha = 3$ (right) on a semi-logarithmic scale. For $\alpha = 3$ ($\alpha = 1$), the width of the traveling wave decreases (increases). For $\alpha = 2$, the width of the traveling wave remains constant.

Since the total size of the population increases super-exponentially on survival and we are mostly interested in long-time behaviour, we set $X(0)$ very large (in the actual implementation, we set $N_1(0) = X(0) = 10^{100}$ and $S_0 = f_1$), which makes fluctuations of the total population size invisible within machine accuracy. Besides, we set $c = 20 \log 10 \approx 46.05$, which gives $p_{k+1}/p_k = 10^{-20}$. Therefore, we have only to consider k up to $\beta S_t X(t) p_k \geq 1$ with $p_k \approx e^{-c(k-1)}$.

Let $\psi_k(t) := N_k(t)/X(t)$. Since parameters are chosen such that deviation from the expected value of $\psi_k(t+1)$ for given $\psi_k(t)$ cannot be generated within machine accuracy, the actual stochastic simulations cannot be different from the deterministic equation

$$\psi_k(t+1) = (1-\beta)\psi_k(t)\frac{f_k}{S_t} + \beta \tilde{p}_k, \quad S_t = \sum_k \psi_k(t) f_k, \tag{79}$$

where $\tilde{p}_k = p_k$ if $\beta X(t+1) p_k > 1$ and 0, otherwise. In a sense, we are studying a deterministic version of the MMM, but, as we mentioned already, even the full stochastic MMM is not distinguishable from the deterministic version MMM for the parameters we chose. Now, we present the numerical solution of (79).

In Figure 1, we depict $\sigma_t \psi(F,t)$ vs $\Delta F/\sigma_t$, where $\Delta F = F - S_t$ on a semi-logarithmic scale at generation $t \approx 10^6$. Here, $\psi(F,t)$ is a density that is calculated as

$$\psi(F,t) = \frac{1}{f_{k+j} - f_{k-j}} \sum_{k-j \leq i \leq k+j} \psi_i(t)$$

with a suitable bin size $2j$, where the integer k is determined uniquely by $f_k \leq F < f_{k+1}$. We assure that dependency of $\psi(F,t)$ on the bin size is negligible over a wide range of j (details not shown here). For comparison, the Gaussian function with zero mean and unit variance is also drawn by a solid curve. Just as we proved for the FMM, the EFD is again well described by a Gaussian traveling wave.

We have found that depending on the actual form of the tail function, σ_t can increase, decrease, or even remain constant in the FMM. To check if this property remains valid in MMM, we plotted the EFD at different times for different values of α, whose result is summarized in Figure 2. The behaviour is the same as shown for the FMM. In fact, the predicted S_t and σ_t for the FMM conform to numerical results (details not shown here). From the numerical observations, we conjecture that the travelling-wave part of the MMM with type I tail function (at least with $n = 1$) has the same EFD as the FMM.

8 Concluding remarks

We provided strong analytical and numerical evidence for the emergence of a travelling wave for the branching process with selection and mutation for unbounded fitness distributions of Gumbel type. For type I tail functions with tail index $n = 1$, or in other words stretched exponential fitness distributions, we show that if the tail parameter satisfies $\alpha > 2$, the standard deviation of the traveling Gaussian wave decreases and eventually the EFD becomes highly peaked like a delta function. Traveling wave solutions of Gaussian form were found previously in a study of the deterministic (infinite population) limit of the model, which amounts to solving the recursion (79) with $\tilde{p}_k = p_k$, see [15]. The expressions for the mean and variance of the EFD obtained in [15] for a particular type I tail function match Eqs. (66) and (67), see also [16].

We conjecture a similar behaviour for bounded fitness distributions of Gumbel type in the condensation case discussed in Section 1. In that case the Gaussian wave is expected to travel to the essential supremum of the fitness distribution, while its standard deviation goes to zero faster than the distance of its mean to the essential supremum. For bounded fitness distributions of Weibull type we conjecture, as in Ref. [9] for a branching model in continuous time, that the condensate emerges in the shape of a Gamma distribution. The conjecture is justified by the rigorous analysis of the deterministic model in Ref. [17].

In our model every individual has a Poisson number of offspring with mean given by its fitness. It is natural to conjecture that results like emergence of the travelling wave, doubly exponential growth rates or condensation also hold for other distributions with the same mean and not too large variance. Verifying this universality conjecture rigorously would be an interesting future project.

Acknowledgements

S.-C.P. acknowledges the support by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant No. RS-2023-00249949) and by The Catholic University of Korea, Research Fund, 2024.

References

[1] Su-Chan Park, Joachim Krug, Léo Touzo, and Peter Mörters. Branching with selection and mutation I: Mutant fitness of Fréchet type. *J. Stat. Phys.*, 190:115, 2023.

[2] Ricardo B.R. Azevedo and Peter Olofsson. A branching process model of evolutionary rescue. *Math. Biosci.*, 341:108708, 2021.

[3] Olivier Couronné and Lucas Gerin. A branching-selection process related to censored Galton–Walton processes. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 50(1):84 – 94, 2014.

[4] Jean Bérard and Jean-Baptiste Gouéré. Brunet-Derrida behavior of branching-selection particle systems on the line. *Commun. Math. Phys.*, 298:323–342, 2010.

[5] Loïc Chaumont and Thi Ngoc Anh Nguyen. On mutations in the branching model for multitype populations. *Adv. Appl. Probab.*, 50(2):543–564, 2018.

[6] Rick Durrett, Jasmine Foo, Kevin Leder, John Mayberry, and Franziska Michor. Evolutionary dynamics of tumor progression with random fitness values. *Theor. Popul. Biol.*, 78:54–66, 2010.

[7] David Cheek and Tibor Antal. Genetic composition of an exponentially growing cell population. *Stoch. Proc. Appl.*, 130:6580–6624, 2020.

[8] Michael D. Nicholson, David Cheek, and Tibor Antal. Sequential mutations in exponentially growing populations. *PLOS Computational Biology*, 19:e1011289, 2023.

[9] Steffen Dereich, Cécile Mailler, and Peter Mörters. Nonextensive condensation in reinforced branching processes. *Ann. Appl. Probab.*, 27:2539 – 2568, 2017.

[10] Laurens de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction.* Springer, Berlin, 2006.

[11] Thomas Bataillon and Susan F. Bailey. Effects of new mutations on fitness: insights from models and data. *Ann. N. Y. Acad. Sci.*, 1320:76–92, 2014.

[12] John H. Gillespie. A simple stochastic gene substitution model. *Theor. Popul. Biol.*, 23:202–215, 1983.

[13] H. Allen Orr. The population genetics of adaptation: The adaptation of DNA sequences. *Evolution*, 56:1317–1330, 2002.

[14] Paul Joyce, Darin R. Rokyta, Craig J. Beisel, and H. Allen Orr. A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. *Genetics*, 180:1627–1643, 2008.

[15] Su-Chan Park and Joachim Krug. Evolution in random fitness landscapes: the infinite sites model. *J. Stat. Mech.: Theory Exp.*, 2008(4):P04014, 2008.

[16] J. F. C. Kingman. A simple model for the balance between selection and mutation. *J. Appl. Probab.*, 15:1–12, 1978.

[17] Steffen Dereich and Peter Mörters. Emergence of condensation in Kingman's model of selection and mutation. *Acta. Appl. Math.*, 127:17–26, 2013.