

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/335989483>

On a model of evolution of subspecies

Preprint · September 2019

DOI: 10.48550/arXiv.1909.09759

CITATIONS

0

READS

28

2 authors:



Rahul Roy

Indian Statistical Institute

58 PUBLICATIONS 1,479 CITATIONS

SEE PROFILE



Hideki Tanemura

Keio University

65 PUBLICATIONS 1,102 CITATIONS

SEE PROFILE

On a model of evolution of subspecies

Rahul Roy and Hideki Tanemura *

Indian Statistical Institute, New Delhi and Keio University, Yokohama

Abstract

Ben-Ari and Schinazi (2016) introduced a stochastic model to study ‘virus-like evolving population with high mutation rate’. This model is a birth and death model with an individual at birth being either a mutant with a random fitness parameter in $[0, 1]$ or having one of the existing fitness parameters with uniform probability; whereas a death event removes the entire population of the least fit site. We change this to incorporate the notion of ‘survival of the fittest’, by requiring that a non-mutant individual, at birth, has a fitness according to a preferential attachment mechanism, i.e., it has a fitness f with a probability proportional to the size of the population of fitness f . Also death just removes one individual at the least fit site. This preferential attachment rule leads to a power law behaviour in the asymptotics, unlike the exponential behaviour obtained by Ben-Ari and Schinazi (2016).

Key words: Markov chain, Random walk, preferential attachment model.

AMS 2000 Subject Classification: 60J10, 60F15, 92D15.

1 Introduction

We study a model of the evolution and survival of species subjected to birth, mutation and death. This model was introduced by Guiol, Machado and Schinazi (2010) and is similar to a model studied by Liggett and Schinazi (2009). It has been of recent interest because of its relation to the discrete evolution model of Bak and Sneppen (1993).

In the model studied by Guiol, Machado and Schinazi (2010), at each discrete time point, with probability p or $1 - p$ respectively, there is either a birth of an individual of the species or a death (in case there exists at least one surviving species). An individual at birth is accompanied by a fitness parameter f , which is chosen uniformly in $[0, 1]$, while the death is always of the individual with the least fitness parameter. They exhibited a phase transition in this model, i.e., for $p > 1/2$,

*E-Mail: rahul@isid.ac.in and tanemura@math.keio.ac.jp

the size of the population, L_n , at time n whose fitness is smaller than $f_c := (1-p)/p$ is a null recurrent Markov chain, while asymptotically, the proportion of the population with fitness level lying in $(a, b) \subseteq (f_c, 1)$ equals $p(b - a)$ almost surely.

In a subsequent paper Ben-Ari and Schinazi (2016) modified the above model to study a ‘virus-like evolving population with high mutation rate’. Here, as earlier, at each discrete time point, with probability p or $1 - p$ respectively, there is either a birth of an individual of the species or a death (in case there exists at least one surviving species) of the individual with the least fitness parameter. The caveat here is that at death, the entire population of the least fit individuals is removed; while, at birth, the individual,

- (i) with probability r , is a mutant and has a fitness parameter f uniformly at random in $[0, 1]$, or
- (ii) with probability $1 - r$, has a fitness parameter chosen uniformly at random among the existing fitness parameters, thereby increasing the population at that fitness level by 1.

For this model too, the authors exhibited a phase transition. In particular, assuming $pr > (1 - p)$, for $f_c := (1 - p)/pr$ the number of fitness levels lying in $(0, f_c)$ at time n where individuals exist is a null recurrent Markov chain, while the number of fitness levels lying to the right of f_c is asymptotically uniformly distributed in $(f_c, 1)$ uniformly.

Here we propose a variant of the Ben-Ari, Schinazi model, a variant which we believe is closer to the Darwinian theory of the survival of the fittest. To incorporate the Darwinian theory, we differ from the above model when a birth occurs which is not a mutant. Instead of the individual at birth having a fitness one of the existing fitness levels chosen uniformly at random, the newly born individual has a fitness f which is chosen proportional to the size of the population of fitness f .

More particularly, suppose that at time n there is a birth, which is not a mutant, and that there are n_i individuals with fitness f_i for $i = 1, \dots, k$ and no other individuals elsewhere. The newly born individual has a fitness f_j with a probability proportional to n_j for $j = 1, \dots, k$. Thus, at birth, an individual without mutation follows a preferential attachment rule akin to the Barabási and Albert (1999) model.

Before we end this section we note that Schreiber (2001) and subsequently Benaïm, Schreiber and Tarrès (2004) study the question of random genetic drift and natural selection via urn models coupled with mean-field behaviour. Unlike our study, there is no spatial aspect of fitness in their model.

A formal set-up of this model is given in the next section, while in the last section we present some mean-field dynamics of the model.

2 The model and statement of results

We first present our model and state the results.

At time 0 there is one individual at site 0. At time n , there is either a birth or a death of an individual from the existing population with probability p or $1 - p$ respectively, where $p \in (0, 1)$, and independent of any other random mechanism considered earlier.

(P1) In case of a birth, there are two possibilities.

- (i) with probability $r \in (0, 1)$, a mutant is born and has a fitness parameter f uniformly at random in $[0, 1]$, or
- (ii) with probability $1 - r$ the individual born has a fitness f with a probability proportional to the number of individuals with fitness f among the entire population present at that time. Here we have a caveat that, if there is no individual present at the time of birth, then the fitness of the individual is sampled uniformly in $[0, 1]$.

(P2) In case of a death, an individual from the population at the site closest to 0 is eliminated.

Here and henceforth, a site represents a fitness level.

Let $X_n = \{(k_i, x_i) : k_i \geq 1, x_i \in [0, 1], i = 1, \dots, l\}$, where the total population at time n is divided in exactly l sites x_1, \dots, x_l , with the size of the population at site x_i being exactly k_i . In case there is no individual present at time n we take $X_n = \emptyset$. The process X_n is Markovian on the state space

$$\mathbb{X} := \{\emptyset\} \cup \{ \{(k, x)\}_{x \in \Lambda} : (k, x) \in \mathbb{N} \times [0, 1], \# \Lambda < \infty, \}, \quad (2.1)$$

where $\mathbb{N} = \{1, 2, \dots\}$.

For a given $f \in (0, 1)$, let L_n^f denote the size of the population at time n at sites in $[0, f]$,

$$L_n^f := \sum_{s \in [0, f]} k_s : s \in [0, f] \text{ and } (k_s, s) \in X_n,$$

R_n^f denote the size of the population at time n at sites in $(f, 1]$,

$$R_n^f := \sum_{s \in (f, 1]} k_s : s \in (f, 1] \text{ and } (k_s, s) \in X_n,$$

and N_n denote the size of the population at time n ,

$$N_n := L_n^f + R_n^f.$$

For a fixed $f \in (0, 1)$, the pair (L_n^f, R_n^f) is a Markov chain on $\mathbb{Z}_+ \times \mathbb{Z}_+$, ($\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$) with transition probabilities given by

(1-1) If $(L_n^f, R_n^f) = (0, 0)$

$$(L_{n+1}^f, R_{n+1}^f) = \begin{cases} (1, 0) & \text{w. p. } fp \\ (0, 1) & \text{w. p. } (1 - f)p \\ (0, 0) & \text{w. p. } 1 - p \end{cases} \quad (2.2)$$

(1-2) If $(L_n^f, R_n^f) \in \{0\} \times \mathbb{N}$

$$(L_{n+1}^f, R_{n+1}^f) = \begin{cases} (1, R_n^f) & \text{w. p. } fpr \\ (0, R_n^f + 1) & \text{w. p. } (1-f)pr + p(1-r) \\ (0, R_n^f - 1) & \text{w. p. } 1-p \end{cases} \quad (2.3)$$

(1-3) If $(L_n^f, R_n^f) \in \mathbb{N} \times \{0\}$

$$(L_{n+1}^f, R_{n+1}^f) = \begin{cases} (L_n^f + 1, 0) & \text{w. p. } fpr + p(1-r) \\ (L_n^f, 1) & \text{w. p. } (1-f)pr \\ (L_n^f - 1, 0) & \text{w. p. } 1-p \end{cases} \quad (2.4)$$

(1-4) If $(L_n^f, R_n^f) \in \mathbb{N} \times \mathbb{N}$

$$(L_{n+1}^f, R_{n+1}^f) = \begin{cases} (L_n^f + 1, R_n^f) & \text{w. p. } fpr + p(1-r)\frac{L_n^f}{N_n} \\ (L_n^f, R_n^f + 1) & \text{w. p. } (1-f)pr + p(1-r)\frac{R_n^f}{N_n} \\ (L_n^f - 1, R_n^f) & \text{w. p. } 1-p. \end{cases} \quad (2.5)$$

The model exhibits a phase transition at a critical position f_c defined as

$$f_c := \frac{1-p}{pr} \quad (2.6)$$

as given in the following theorem:

Theorem 1 (1) *In case $p \leq 1-p$, the population dies out infinitely often a.s., in the sense that*

$$P(N_n = 0 \text{ for infinitely many } n) = 1 \quad (2.7)$$

(2) *In case $1-p < rp$, the size of the population goes to infinity as $n \rightarrow \infty$, and most of the population is distributed at sites in the interval $[f_c, 1]$, in the sense that*

$$P(\lim_{n \rightarrow \infty} \frac{R_n^{f_c}}{N_n} = 1) = 1 \text{ and } P(\liminf_{n \rightarrow \infty} \frac{R_n^{f_c} - R_n^f}{N_n} > 0) = 1 \text{ for any } f > f_c. \quad (2.8)$$

(3) *In case $rp \leq 1-p < p$, the size of the population goes to infinity as $n \rightarrow \infty$, and most of the population is concentrated at sites near 1, in the sense that*

$$P(\lim_{n \rightarrow \infty} N_n = \infty) = 1 \text{ and, for any } \varepsilon > 0, P(\lim_{n \rightarrow \infty} \frac{R_n^{1-\varepsilon}}{N_n} = 1) = 1. \quad (2.9)$$

Let $F_n(f)$ denote the empirical distribution of sites at time n , i.e.

$$F_n(f) := \frac{\#\{s \in [0, f] : (k, s) \in X_n \text{ for some } k \geq 1\}}{\#\{s \in [0, 1] : (k, s) \in X_n \text{ for some } k \geq 1\}},$$

we have

Corollary 2 *If $1 - p < rp$ (i.e., $f_c < 1$), then*

$$F_n(f) \rightarrow \frac{\max\{f - f_c, 0\}}{1 - f_c} \quad \text{uniformly a.s.} \quad (2.10)$$

Let $S_n := \#\{s \in [0, 1] : (k, s) \in X_n \text{ for some } k \geq 1\}$ be the total number of sites at time n among which the total population is distributed. For a given n, k, f let $U_n^k(f) := \#\{s \in [f, 1] : (k, s) \in X_n\}$ denote the number of sites in $[f, 1]$ at time n which has a population of size exactly k ; clearly $S_n = \sum_k U_n^k(0)$. Taking $U_n^k(f+) = \lim_{s \downarrow f} U_n^k(s)$, for $A \subseteq \mathbb{X}$, define the empirical distribution of size and fitness on $\mathbb{N} \times [0, 1]$ as

$$H_n(A) := \begin{cases} \frac{\sum_{(k,f) \in A} U_n^k(f) - U_n^k(f+)}{S_n}, & S_n > 0, \\ \delta_{(0,0)}(A), & S_n = 0. \end{cases} \quad (2.11)$$

Theorem 3 *For $pr > 1 - p$, as $n \rightarrow \infty$, H_n converges weakly to a product measure on $\mathbb{N} \times [0, 1]$ whose density is given by*

$$p_k \frac{\mathbf{1}_{[f_c, 1]}(x)}{1 - f_c} dx, \quad (k, x) \in \mathbb{N} \times [0, 1]$$

$$\text{with } p_k = \frac{(2p-1)r}{(1-r)(1-p)} B\left(1 + \frac{(2p-1)r}{(1-r)(1-p)}, k\right) \text{ for } k \in \mathbb{N}, \quad (2.12)$$

where $B(a, b)$ is the Beta function with parameter $a, b > 0$.

Remark 1 *Since $B(s, k) = \mathcal{O}(k^{-s})$, $k \rightarrow \infty$, the probability density p_k , $k \in \mathbb{N}$ has m -th moment if and only if $r > 1 - \frac{2p-1}{2p-1+(1-p)m}$.*

For the model studied by Ben-Ari and Schinazi (2016), in case of a death, the entire population at the site of lowest fitness is removed unlike our condition (P2). Thus in their model, if \tilde{S}_n denotes the number of sites at time n among which the total population is distributed, then \tilde{S}_n is a Markov chain with spatially homogeneous transition probabilities given by

$$\tilde{S}_{n+1} = \begin{cases} \tilde{S}_n + 1 & \text{with probability } pr, \\ \tilde{S}_n & \text{with probability } p(1-r), \\ \tilde{S}_n - 1 & \text{with probability } 1-p, \end{cases} \quad (2.13)$$

with reflecting boundary condition at 0. For a given $f \in (0, 1)$, letting \tilde{S}_n^{f-} denote the number of sites at time n in $[0, f]$, and \tilde{S}_n^{f+} the number of sites at the sites in $(f, 1]$, the pair $(\tilde{S}_n^{f-}, \tilde{S}_n^{f+})$ is a spatially homogeneous Markov chain on $\mathbb{Z}_+ \times \mathbb{Z}_+$, where $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$:

(BAS-1) If $(\tilde{S}_n^{f-}, \tilde{S}_n^{f+}) = (0, 0)$

$$(\tilde{S}_{n+1}^{f-}, \tilde{S}_{n+1}^{f+}) = \begin{cases} (1, 0) & \text{w. p. } fp \\ (0, 1) & \text{w. p. } (1-f)p \\ (0, 0) & \text{w. p. } 1-p \end{cases} \quad (2.14)$$

(BAS-2) If $(\tilde{S}_n^{f-}, \tilde{S}_n^{f+}) \in \{0\} \times \mathbb{N}$

$$(\tilde{S}_{n+1}^{f-}, \tilde{S}_{n+1}^{f+}) = \begin{cases} (1, \tilde{S}_n^{f+}) & \text{w. p. } fpr \\ (0, \tilde{S}_n^{f+} + 1) & \text{w. p. } (1-f)pr \\ (0, \tilde{S}_n^{f+}) & \text{w. p. } p(1-r) \\ (0, \tilde{S}_n^{f+} - 1) & \text{w. p. } 1-p \end{cases} \quad (2.15)$$

(BAS-3) If $(\tilde{S}_n^{f-}, \tilde{S}_n^{f+}) \in \mathbb{N} \times \{0\}$

$$(\tilde{S}_{n+1}^{f-}, \tilde{S}_{n+1}^{f+}) = \begin{cases} (\tilde{S}_n^{f-} + 1, 0) & \text{w. p. } fpr \\ (\tilde{S}_n^{f-}, 1) & \text{w. p. } (1-f)pr \\ (\tilde{S}_n^{f-}, 0) & \text{w. p. } p(1-r) \\ (\tilde{S}_n^{f-} - 1, 0) & \text{w. p. } 1-p \end{cases} \quad (2.16)$$

(BAS-4) If $(\tilde{S}_n^{f-}, \tilde{S}_n^{f+}) \in \mathbb{N} \times \mathbb{N}$

$$(\tilde{S}_{n+1}^{f-}, \tilde{S}_{n+1}^{f+}) = \begin{cases} (\tilde{S}_n^{f-} + 1, \tilde{S}_n^{f+}) & \text{w. p. } fpr \\ (\tilde{S}_n^{f-}, \tilde{S}_n^{f+}) & \text{w. p. } p(1-r) \\ (\tilde{S}_n^{f-}, \tilde{S}_n^{f+} + 1) & \text{w. p. } (1-f)pr \\ (\tilde{S}_n^{f-} - 1, \tilde{S}_n^{f+}) & \text{w. p. } 1-p \end{cases} \quad (2.17)$$

Also at birth, if the individual is not a mutant then the individual born has a fitness chosen uniformly at random among the fitnesses of the existing individuals at that time, unlike the preferential condition (P1)(ii) of our model. As such, the transition probabilities for this model are spatially homogeneous, while for our model, as is exemplified by (2.5), the transition probabilities are not spatially homogeneous. Thus the equivalent result they have for Theorem 3 has p_k arising from a Geom $\left(\frac{pr-(1-p)}{p-(1-p)}\right)$ distribution.

The power law phenomenon present in the study of preferential attachment graphs (see van der Hofstad (2017) Chapter 8) manifests itself in our model (as noted in Remark 1) through the Beta function in Theorem 3.

3 Proof of Theorem 1

As noted in Guiol, Machado and Schinazi (2010), for $p \leq 1 - p$, i.e. when the death rate is more than the birth rate, the process $\{N_n : n \geq 0\}$ is equivalent to a random walk on the non-negative integers \mathbb{Z}_+ with non-positive drift and a holding at 0 with probability $(1 - p)$. Thus N_n returns to the 0 infinitely often with probability 1.

For $p > 1 - p$, $\{N_n : n \geq 0\}$ is equivalent to a random walk on the non-negative integers \mathbb{Z}_+ with positive drift and thus $N_n \rightarrow \infty$ as $n \rightarrow \infty$ with probability 1.

Then we study the case when $1 - p < p$.

Lemma 4 (1) Let $f_c = \frac{1-p}{rp} < 1$.

(i) For $f < f_c$ and for any $\eta \in (0, 1)$ we have

$$P\left(\text{there exists } T > 0 \text{ such that } \rho_n^f \equiv \frac{L_n^f}{N_n} \leq \eta \text{ for all } n \geq T\right) = 1, \quad (3.1)$$

and

$$P(L_n^f = 0 \text{ infinitely often}) = 1. \quad (3.2)$$

(ii) Let $f > f_c$. Then

$$P(L_n^f = 0 \text{ infinitely often}) = 0. \quad (3.3)$$

(2) Let $1 \leq f_c = \frac{1-p}{rp} < \frac{1}{r}$.

(i) For $f < 1$ and for any $\eta \in (0, 1)$ we have (3.1) and (3.2).

(ii) Let $f = 1$. Then we have (3.3).

Proof. We prove two cases (1) and (2) together. The idea of the proof is that, since for $f < f_c \wedge 1$, R_n^f will be much larger than L_n^f , we stochastically bound the non-spatially homogeneous Markov chain with a boundary condition by a spatially homogeneous Markov chain a boundary condition, and study the modified Markov chain. As such, for $\varepsilon \in [0, 1]$, we introduce a Markov chain $(L_n^f(\varepsilon), R_n^f(\varepsilon))$ with stationary transition probabilities given by

(Ep-1) If $(L_n^f(\varepsilon), R_n^f(\varepsilon)) = (0, 0)$

$$(L_{n+1}^f(\varepsilon), R_{n+1}^f(\varepsilon)) = \begin{cases} (1, 0) & \text{w. p. } fp \\ (0, 1) & \text{w. p. } (1-f)p \\ (0, 0) & \text{w. p. } 1-p. \end{cases} \quad (3.4)$$

(Ep-2) If $(L_n^f(\varepsilon), R_n^f(\varepsilon)) \in \{0\} \times \mathbb{N}$

$$(L_{n+1}^f(\varepsilon), R_{n+1}^f(\varepsilon)) = \begin{cases} (1, R_n^f(\varepsilon)) & \text{w. p. } fpr \\ (0, R_n^f(\varepsilon) + 1) & \text{w. p. } (1-f)pr + p(1-r) \\ (0, R_n^f(\varepsilon) - 1) & \text{w. p. } 1-p. \end{cases} \quad (3.5)$$

(Ep-3) If $(L_n^f(\varepsilon), R_n^f(\varepsilon)) \in \mathbb{N} \times \{0\}$

$$(L_{n+1}^f(\varepsilon), R_{n+1}^f(\varepsilon)) = \begin{cases} (L_n^f(\varepsilon) + 1, 0) & \text{w. p. } fpr + p(1-r) \\ (L_n^f(\varepsilon), 1) & \text{w. p. } (1-f)pr \\ (L_n^f(\varepsilon) - 1, 0) & \text{w. p. } 1-p. \end{cases} \quad (3.6)$$

(Ep-4) If $(L_n^f(\varepsilon), R_n^f(\varepsilon)) \in \mathbb{N} \times \mathbb{N}$

$$(L_{n+1}^f(\varepsilon), R_{n+1}^f(\varepsilon)) = \begin{cases} (L_n^f(\varepsilon) + 1, R_n^f(\varepsilon)) & \text{w. p. } fpr + p(1-r)\varepsilon \\ (L_n^f(\varepsilon), R_n^f(\varepsilon) + 1) & \text{w. p. } (1-f)pr + p(1-r)(1-\varepsilon) \\ (L_n^f(\varepsilon) - 1, R_n^f(\varepsilon)) & \text{w. p. } 1-p. \end{cases} \quad (3.7)$$

For $\varepsilon \in [0, 1]$, we couple the processes $\{(L_n^f(\varepsilon), R_n^f(\varepsilon)) : n \geq 1\}$ such that

$$L_n^f(\varepsilon) \leq L_n^f(\varepsilon'), \quad R_n^f(\varepsilon) \geq R_n^f(\varepsilon') \quad \text{for } \varepsilon \leq \varepsilon' \text{ and all } n \geq 1. \quad (3.8)$$

Taking L_n^f , R_n^f and N_n as in Subsection 2.1 and $L_n^f(\cdot)$ and $R_n^f(\cdot)$ as above, we have, for $\rho_n^f := \frac{L_n^f}{N_n}$,

$$N_n(\varepsilon) := L_n^f(\varepsilon) + R_n^f(\varepsilon) = N_n \quad (3.9)$$

$$L_{n+1}^f = L_{n+1}^f(\rho_n^f), \quad R_{n+1}^f = R_{n+1}^f(\rho_n^f), \quad (3.10)$$

$$L_n^f(0) \leq L_n^f \leq L_n^f(1), \quad R_n^f(1) \leq R_n^f \leq R_n^f(0). \quad (3.11)$$

By the law of large numbers we have

$$\lim_{n \rightarrow \infty} \frac{L_n^f(\varepsilon)}{n} = [fpr + p(1-r)\varepsilon - 1 + p]_+, \text{ and } \lim_{n \rightarrow \infty} \frac{N_n}{n} = 2p - 1, \text{ almost surely,}$$

and so, for $\rho_n^f(\varepsilon) := \frac{L_n^f(\varepsilon)}{N_n}$, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \rho_n^f(\varepsilon) &= \left[\frac{fpr + p(1-r)\varepsilon - 1 + p}{2p - 1} \right]_+ \\ &= \left[\frac{fpr - 1 + p}{2p - 1} + \frac{p(1-r)\varepsilon}{2p - 1} \right]_+. \end{aligned} \quad (3.12)$$

We introduce the linear function defined by

$$h(x) = \frac{fpr - 1 + p}{2p - 1} + \frac{p(1-r)}{2p - 1}x.$$

Note that $\frac{p(1-r)}{2p-1} > 0$. By a simple calculation we see that if $f \leq 1$

$$h(0) \leq \frac{pr - 1 + p}{2p - 1} < 0 \quad \text{if } pr < 1 - p \quad \text{and} \quad h(1) \leq 1 - \frac{2p}{2p - 1} < 1.$$

Then we may choose $\delta > 0$ such that

$$h_\delta(x) := h(x + \delta) < x, \quad [0, 1]. \quad (3.13)$$

Put

$$\Lambda(\varepsilon, \delta) = \left\{ \omega : \text{there exists } N = N(\omega) \in \mathbb{N} \text{ such that for all } n \geq N, \rho_n^f(\varepsilon) < h_\delta(\varepsilon) \right\}.$$

From (3.12), we have that

$$P(\Lambda(\varepsilon, \delta)) = 1, \quad \text{for all } \varepsilon, \delta \in (0, 1]. \quad (3.14)$$

Also, taking $\varepsilon_c > 0$ such that $h_\delta(\varepsilon_c) = \eta$, i.e.,

$$\varepsilon_c = \varepsilon_c(\delta, \eta) := \frac{\eta - \frac{fpr-1+p}{2^{p-1}}}{\frac{p(1-r)}{2^{p-1}}(1+\delta)} = \frac{(2p-1)\eta - (fpr-1+p)}{p(1-r)(1+\delta)},$$

we see that for $\varepsilon \leq \varepsilon_c$ we have $\max\{h_\delta(\varepsilon), \eta\} = \eta$.

Now consider the recursion formula

$$x_{n+1} = h_\delta(x_n). \quad (3.15)$$

Since (3.13), for $f < f_c < 1$,

$$x_n \text{ is decreasing and } \lim_{n \rightarrow \infty} x_n = \frac{fpr-1+p}{pr-1+p} < 0. \quad (3.16)$$

We put

$$h_\delta(k, x) := h_\delta(2^{-k}([2^k x] + 1)) \text{ for } k \in \mathbb{N},$$

where $[a]$ the largest integer less than $a \in \mathbb{R}$. From (3.16) we see that, for sufficient large k , there exists $n_c \in \mathbb{N}$ such that

$$h_\delta^n(k, 1) := h_\delta(k, h_\delta^{n-1}(k, 1)) \leq \eta \text{ for all } n \geq n_c. \quad (3.17)$$

Note that from (3.8) and (3.11) we have that

$$\rho_n^f(\varepsilon) \geq \rho_n^f(\varepsilon') \text{ for } \varepsilon > \varepsilon', \text{ and } \rho_n^f \leq \rho_n^f(1), \quad (3.18)$$

thus, for any $\omega \in \bigcap_{m \in \mathbb{N}} \Lambda(m2^{-k}, \delta)$ there exists $N_1(\omega)$ such that, for all $n \geq N_1(\omega)$,

$$\rho_n^f[\omega] \leq \rho_n^f(1)[\omega] \leq h_\delta(k, 1),$$

and there exists $N_2(\omega) \geq N_1(\omega)$ such that for all $n \geq N_2(\omega)$

$$\rho_n^f[\omega] \leq \rho_n^f(h_\delta(k, 1))[\omega] \leq h_\delta(k, h(k, 1)) = h_\delta^2(k, 1).$$

Repeating this procedure we have for any $\ell \in \mathbb{N}$ there exists $N_\ell(\omega)$ such that for all $n \geq N_\ell(\omega)$

$$\rho_n^f[\omega] \leq h_\delta^\ell(k, 1). \quad (3.19)$$

From (3.17), we now have

$$\rho_n^f[\omega] \leq \eta \quad \text{for all } n \geq N_{n_c}(\omega).$$

Since $P(\bigcap_{m \in \mathbb{N}} \Lambda(m2^{-k}, \delta)) = 1$ from (3.14), we have

$$\lim_{n \rightarrow \infty} \rho_n^f[\omega] \leq \eta, \quad \text{a.s.} \quad (3.20)$$

Thus we obtain (3.1).

If

$$f < f_c - \frac{1-r}{r}\varepsilon, \quad (3.21)$$

$L_n^f(\varepsilon)$ is recurrent. Also, for $f < f_c \wedge 1$, the condition (3.21) holds for sufficiently small ε , hence from (3.1) we see that L_n^f hits the origin infinitely often. This proves (i) of the Lemma 4.

Let $f_c < 1$. Observing that, for \tilde{S}_n^{f-} as in (2.13) and $L_n^f(\cdot)$ as above,

$$\tilde{S}_n^{f-} \leq L_n^f(0),$$

we see from (2.14)-(2.17) that when $f > f_c$, for only finitely many n we have $\tilde{S}_n^{f-} = 0$. Thus, from (3.11) we have (ii) of (1). Let $1 \leq f_c < \frac{1}{r}$. Since $1-p < p$ the random walk comparison as noted at the beginning of this section shows that $N_n \rightarrow \infty$ almost surely as $n \rightarrow \infty$. We have (ii) of (2). ■

We give the proof of Theorem 1. Part (1) is obtained by the random walk comparison. Part (3) is derived from (2) of Lemma 4. The first statement of (2) is derived (ii) of (1) and (2) in Lemma 4.

Finally, considering the birth rate rp of mutants, the limiting expected number of them with a fitness between (a, b) , with $f_c < a < b \leq 1$, is $rp(b-a)$. Thus we have, by an application of the strong law of large numbers

$$\liminf_{n \rightarrow \infty} \frac{R_n^b - R_n^a}{N_n} \geq \frac{p(b-a)}{2p-1} \text{ almost surely.}$$

(Note this also follows from part (b) of the main Theorem of Guiol, Machado and Schinazi (2010).) This completes the proof of the second statement of part (2) of Theorem 1.

Finally, since the sites are each independently and uniformly distributed on $[0, 1]$ Corollary 2 follows from Lemma 4.

4 Proof of Theorem 3

We will prove Theorem 3 with the help of two lemmas.

Let $A_k(t_1, n)$, $k, t_1, n \in \mathbb{N}$, be the event that a mutant born at time t_1 gets $k - 1$ attachments until time n , and let $q_k(t_1, n) := P(A_k(t_1, n))$. We have

Lemma 5 *Let $p = 1$ i.e. no deaths. For each $k, t_1 \in \mathbb{N}$*

$$E \left[\left\{ \frac{1}{n} \sum_{t_1=1}^n (\mathbf{1}_{A_k(t_1, n)} - q_k(t_1, n)) \right\}^2 \right] \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (4.1)$$

Proof. The left hand side of (4.1) is

$$\begin{aligned} & \frac{1}{n^2} \sum_{t_1=1}^n \sum_{s_1=1}^n [P(A_k(s_1, n) \cap A_k(t_1, n)) - P(A_k(s_1, n))P(A_k(t_1, n))] \\ &= \frac{1}{n^2} \sum_{t_1=1}^n \sum_{s_1=1}^n P(A_k(s_1, n)) [P(A_k(t_1, n) | A_k(s_1, n)) - P(A_k(t_1, n))]. \end{aligned}$$

Thus it is enough to show the following for the proof of the lemma: for any $x_1, y_1 \in (0, 1)$ with $x_1 < y_1$

$$P(A_k(y_1 n, n) | A_k(x_1 n, n)) - P(A_k(y_1 n, n)) \rightarrow 0, \quad n \rightarrow \infty. \quad (4.2)$$

Let $\{t_\ell\}_{\ell=1}^k$ be an increasing sequence of \mathbb{N} with $t_k \leq n$. We denote by $A_k[\{t_\ell\}_{\ell=1}^k; n]$ the event that a mutant comes at time t_1 which gets its $(\ell - 1)$ th attachment at time t_ℓ , $\ell = 2, 3, \dots, k$, and no other attachment till time n . Then

$$A_k(t_1, n) = \sum_{\substack{t_2, t_3, \dots, t_k \in \mathbb{N} \\ t_1 < t_2 < \dots < t_k < n}} A_k[\{t_\ell\}_{\ell=1}^k; n]. \quad (4.3)$$

Let $\{s_\ell\}_{\ell=1}^k$ and $\{t_\ell\}_{\ell=1}^k$ be increasing sequences of \mathbb{N} with $s_k, t_k \leq n$.

Suppose that $s_1 = t_1$, then

$$P(A_k[\{t_\ell\}_{\ell=1}^k; n] | A_k[\{s_\ell\}_{\ell=1}^k; n]) = \mathbf{1}(s_\ell = t_\ell, \ell = 2, 3, \dots, k). \quad (4.4)$$

Also, for $s_1 \neq t_1$, if $\{s_\ell; \ell = 2, \dots, k\} \cap \{t_\ell; \ell = 2, \dots, k\} \neq \emptyset$, then

$$P(A_k[\{t_\ell\}_{\ell=1}^k; n] | A_k[\{s_\ell\}_{\ell=1}^k; n]) = 0; \quad (4.5)$$

and if $\{s_\ell; \ell = 1, 2, \dots, k\} \cap \{t_\ell; \ell = 1, 2, \dots, k\} = \emptyset$, then

$$\begin{aligned} & P(A_k[\{t_\ell\}_{\ell=1}^k; n] | A_k[\{s_\ell\}_{\ell=1}^k; n]) \\ &= P(A_k[\{t_\ell\}_{\ell=1}^k; n] | \text{the mutant which came at time } t_1 \\ & \quad \text{does not get any attachment at times } \{s_\ell\}_{\ell=1}^k) \\ &= P(A_k[\{t_\ell\}_{\ell=1}^k; n]) \prod_{m: s_m > t_1} \left(1 - \frac{\ell[s_m](1-r)}{s_m} \right)^{-1}, \end{aligned}$$

where $\ell[s_m] = \max\{\ell : t_\ell < s_m\}$ is the population size at time s_m of the fitness location occupied by the mutant which came at time t_1 . Hence, we have,

$$\begin{aligned} & P(A_k[\{t_\ell\}_{\ell=1}^k; n] | A_k[\{s_\ell\}_{\ell=1}^k; n]) - P(A_k[\{t_\ell\}_{\ell=1}^k; n]) \\ &= P(A_k[\{t_\ell\}_{\ell=1}^k; n] | A_k[\{s_\ell\}_{\ell=1}^k; n]) \left[1 - \prod_{m: s_m > t_1} \left(1 - \frac{\ell[s_m](1-r)}{s_m} \right) \right] \\ &\leq \frac{k^2}{t_1} P(A_k[\{t_\ell\}_{\ell=1}^k; n] | A_k[\{s_\ell\}_{\ell=1}^k; n]). \end{aligned} \quad (4.6)$$

Combining (4.4), (4.5) and (4.6) with (4.3), we obtain (4.2). This completes the proof. ■

Next we have

Lemma 6 *Let $p = 1$. For each $k \in \mathbb{N}$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t_1=1}^n q_k(t_1, n) = \frac{r}{1-r} B\left(\frac{2-r}{1-r}, k\right) = p_k. \quad (4.7)$$

Proof. Let $A_k(t_1, n)$ and $q_k(t_1, n)$, $k, t_1, n \in \mathbb{N}$, be as above. For $k = 1$, we have

$$q_1(t_1, n) = r \prod_{j=t_1+1}^n \left(1 - \frac{1-r}{j} \right),$$

since the number of individuals at time $j-1$ is j and the probability that the mutant who arrived at time t_1 gets an attachment at time j is $\frac{1-r}{j}$.

For $k = 2$

$$q_2(t_1, n) = r \sum_{t_2=t_1+1}^n \left\{ \prod_{j=t_1+1}^{t_2-1} \left(1 - \frac{1-r}{j} \right) \right\} \frac{1-r}{t_2} \left\{ \prod_{j=t_2+1}^n \left(1 - \frac{2(1-r)}{j} \right) \right\},$$

where t_2 is the time of the first attachment. Similarly for each $k \in \mathbb{N}$

$$q_k(t_1, n) = r \sum_{t_1 < t_2 < \dots < t_k \leq n} \prod_{\ell=1}^k \prod_{j=t_{\ell-1}+1}^{t_\ell} \left(1 - \frac{\ell(1-r)}{j} \right) \prod_{\ell=1}^{k-1} \frac{\ell(1-r)}{t_{\ell+1} - \ell(1-r)},$$

where we used the equation

$$\frac{\ell(1-r)}{t_{\ell+1}} \frac{1}{1 - \frac{\ell(1-r)}{t_{\ell+1}}} = \frac{\ell(1-r)}{t_{\ell+1} - \ell(1-r)}.$$

By using Stirling's formula we see that

$$\prod_{j=t_\ell+1}^{t_{\ell+1}} \left(1 - \frac{\ell(1-r)}{j} \right) \sim \left(\frac{t_\ell}{t_{\ell+1}} \right)^{\ell(1-r)}, \quad t_\ell, t_{\ell+1} \rightarrow \infty.$$

Now letting $n \rightarrow \infty$ and taking $t_\ell = nx_\ell$ we have

$$\begin{aligned}
\frac{1}{n} \sum_{t_1=1}^n q_k(t_1, n) &\sim r \int_{0 < x_1 < \dots < x_k < 1} dx_1 \cdots dx_k \prod_{\ell=1}^k \left(\frac{x_\ell}{x_{\ell+1}} \right)^{\ell(1-r)} \prod_{\ell=1}^{k-1} \frac{\ell(1-r)}{x_{\ell+1}} \\
&= r(1-r)^{k-1} (k-1)! \int_{0 < x_1 < \dots < x_k < 1} dx_1 \cdots dx_k x_1^{1-r} \prod_{\ell=2}^k x_\ell^{-r} \\
&= r(1-r)^{k-1} \int_0^1 dx_1 x_1^{1-r} \prod_{\ell=2}^k \int_{x_1}^1 dx_\ell x_\ell^{-r} \\
&= r \int_0^1 dx_1 x_1^{1-r} (1 - x_1^{1-r})^{k-1} \\
&= \frac{r}{1-r} \int_0^1 dy y^{\frac{1}{1-r}} (1-y)^{k-1} = \frac{r}{1-r} B\left(\frac{2-r}{1-r}, k\right).
\end{aligned}$$

This completes the proof. ■

We give the proof of Theorem 3. When $p = 1$ From Lemmas 5 and 6 we have

$$\frac{1}{n} \sum_{t_1=1}^n \mathbf{1}_{A_k(t_1, n)} \rightarrow \frac{r}{1-r} B\left(\frac{2-r}{1-r}, k\right) \text{ as } n \rightarrow \infty, \quad \text{in probability.}$$

Noting that

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = r, \quad \text{a.s.}$$

we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{f \in (0,1)} U_n^k(f) - U_n^k(f+)}{S_n} = \frac{1}{1-r} B\left(\frac{2-r}{1-r}, k\right) = p_k. \quad \text{in probability.} \quad (4.8)$$

Next we consider the case where $p \in (0, 1)$. We introduce another Markov process \hat{X}_n , $n \in \mathbb{N} \cup \{0\}$, which is a pure birth process, as follows:

1. At time 0 there exists one individual at a site uniformly distributed on $(f_c, 1)$.
2. with probability $p(1 - rf_c)$ there is a new birth. There are two possibilities –
 - with probability $\hat{r} := \frac{pr(1 - f_c)}{p(1 - rf_c)}$ a mutant is born with a fitness uniformly distributed in $[f_c, 1]$,
 - with probability $1 - \hat{r} := \frac{p(1 - r)}{p(1 - rf_c)}$ a non-mutant individual is born. It has a fitness f with a probability proportional to the number of individuals of fitness f , and we increase the corresponding population of fitness f individuals by 1.

3. With probability $1 - p(1 - rf_c)$ nothing happens, i.e. neither a birth nor a death occurs.

For the Markov process \hat{X}_n , $n \in \mathbb{N} \cup \{0\}$, we define \hat{q}_k , \hat{S}_n and \hat{U}_n in the same manner as q_k , S_n and U_n for X_n , $n \in \mathbb{N} \cup \{0\}$. Then by the same argument as above we see that

$$\frac{1}{n} \sum_{t_1=1}^n \tilde{q}_k(t_1, n) \sim p(1 - rf_c) \frac{\hat{r}}{1 - \hat{r}} B\left(\frac{2 - \hat{r}}{1 - \hat{r}}, k\right)$$

and

$$\lim_{n \rightarrow \infty} \frac{\hat{S}_n}{n} = pr(1 - f_c).$$

Hence

$$\lim_{n \rightarrow \infty} \frac{\sum_{f \in (0,1)} \hat{U}_n^k(f) - \hat{U}_n^k(f+)}{\hat{S}_n} = \frac{1}{1 - \hat{r}} B\left(\frac{2 - \hat{r}}{1 - \hat{r}}, k\right) = p_k$$

From Lemma 4, we know that deletions of individuals in $(f_c, 1)$ occur finitely often and $\frac{R_n^f}{L_n^f + R_n^f} \rightarrow 1$ almost surely as $n \rightarrow \infty$. Thus we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{f \in (0,1)} U_n^k(f) - U_n^k(f+)}{S_n} = \lim_{n \rightarrow \infty} \frac{\sum_{f \in (0,1)} \hat{U}_n^k(f) - \hat{U}_n^k(f+)}{\hat{S}_n} \quad \text{a.s.}$$

and so (4.8) for $p \in (0, 1]$. Noting that the sites are uniformly distributed on $[0, 1]$ independently, and preferential attachment does not depend on the position of sites, we obtain Theorem 3 from (4.8). ■

5 Number of individuals of a fixed fitness

Fix $f \in [0, 1]$ and let N_n^f denote the number of individuals with fitness f at time n . When $rp > 1 - p$, i.e. $f_c < 1$, from Lemma 4 we know that, $P(L_n^f = 0 \text{ infinitely often}) = 1$ for $f \in (f_c, 1)$. Thus, if a mutant with fitness $f \in (f_c, 1)$ is born at some large time ℓ , then the chances of the mutant dying is small, and so a natural question is ‘for some $n > \ell$, how many individuals did this mutant attract by time n ’, i.e., what is the value of N_n^f ?

Proposition 7 *Fix $f \in (f_c, 1)$, we have, for $\ell < n$, as $\ell, n \rightarrow \infty$*

$$\begin{aligned} & E[N_n^f | \text{a mutant with fitness } f \text{ is born at time } \ell] \\ & \sim \frac{\Gamma((2p - 1)\ell + 1) \Gamma((2p - 1)n + 1 + p(1 - r))}{\Gamma((2p - 1)\ell + 1 + p(1 - r)) \Gamma((2p - 1)n + 1)} \\ & \sim \left(\frac{n}{\ell}\right)^{p(1-r)}. \end{aligned}$$

Proof. Since we are interested in the region $f > f_c$ and also, for the calculation of the expectation, we just need to factor out the death rate $(1 - p)$, so we modify the Markov process \hat{X}_n introduced in the proof of Lemma 6, by removing the times when ‘nothing happens’, i.e. the process does not move. This is done as follows: let \hat{N}_n be the number of individuals of the process \hat{X}_n at time n , we define a new Markov process \check{X}_n , for $n \geq 0$, by

$$\hat{X}_n = \check{X}_{\hat{N}_n-1}.$$

Since $\hat{N}_0 = 1$, we see that $\check{N}_\ell = \ell + 1$, where \check{N}_ℓ is the number of individuals of the process \check{X} at time ℓ .

Letting \check{N}_m^f denote the number of individuals of the \check{X} process of fitness f at time m , we have

$$\begin{aligned} E[\check{N}_m^f | \check{N}_{m-1}^f] &= \{1 - p(1 - r)\} \check{N}_{m-1}^f + p(1 - r) \left\{ (\check{N}_{m-1}^f + 1) \frac{\check{N}_{m-1}^f}{m} + \check{N}_{m-1}^f \left(1 - \frac{\check{N}_{m-1}^f}{m} \right) \right\} \\ &= \left(1 + \frac{p(1 - r)}{m} \right) \check{N}_{m-1}^f. \end{aligned}$$

If $\check{N}_0^f = \check{N}_0 = 1$ then we have

$$E[\check{N}_m^f | \check{N}_0^f = 1] = \prod_{k=1}^m \left(\frac{k + p(1 - r)}{k} \right) = \frac{\Gamma(m + 1 + p(1 - r))}{\Gamma(1 + p(1 - r))\Gamma(m + 1)}, \quad (5.1)$$

while, if $\check{N}_\ell^f = 1$ then we have

$$E[\check{N}_m^f | \check{N}_\ell^f = 1] = \prod_{k=\ell+1}^m \left(\frac{k + p(1 - r)}{k} \right) = \frac{\Gamma(\ell + 1)\Gamma(m + 1 + p(1 - r))}{\Gamma(\ell + 1 + p(1 - r))\Gamma(m + 1)}. \quad (5.2)$$

Since $\frac{\hat{N}_n}{n} \rightarrow pr(1 - f_c) + p(1 - r) = 2p - 1$, if $\hat{N}_0^f = 1$ then we have

$$E[\hat{N}_n^f | \hat{N}_0^f = 1] \sim \prod_{k=1}^{(2p-1)n} \left(\frac{k + p(1 - r)}{k} \right) = \frac{\Gamma((2p - 1)n + 1 + p(1 - r))}{\Gamma(1 + p(1 - r))\Gamma((2p - 1)n + 1)}. \quad (5.3)$$

Also, $\frac{\hat{N}_\ell}{\ell} \rightarrow 2p - 1$, so for $\hat{N}_\ell^f = 1$, we have

$$\begin{aligned} E[\hat{N}_n^f | \hat{N}_\ell^f = 1] &= \prod_{k=(2p-1)\ell+1}^{(2p-1)n} \left(\frac{k + p(1 - r)}{k} \right) \\ &= \frac{\Gamma((2p - 1)\ell + 1)\Gamma((2p - 1)n + 1 + p(1 - r))}{\Gamma((2p - 1)\ell + 1 + p(1 - r))\Gamma((2p - 1)n + 1)}. \end{aligned}$$

From Lemma 4 we have $E[N_n^f | N_\ell^f = 1] \sim E[\hat{N}_n^f | \hat{N}_\ell^f = 1]$, and that completes the proof of the proposition. ■

6 Heuristics for the case $f_c > 1$

We now present some mean field heuristics about the location of the leftmost site x_t at time t in the case when $pr < 1 - p < p$, i.e. $f_c > 1$. These heuristics should be seen in connection with (2.9) of Theorem 1.

Let $y_t = 1 - x_t$. The number of individuals to enter the interval $(x_t, 1]$ is approximately

$$pr y_t dt + p(1 - r)dt,$$

where the first term counts the births which are mutants and the second term counts the births which are not mutants. While the number of individuals deleted in the interval $(x_t, 1]$ is approximately

$$-\frac{dy_t}{y_t}\{p - (1 - p)\}t,$$

this being the absolute value of the deletions since $\frac{dy_t}{dt} < 0$. Thus we consider the following differential equation:

$$pr y_t dt + p(1 - r)dt + \frac{dy_t}{y_t}\{2p - 1\}t = (2p - 1)dt,$$

from which we have

$$\begin{aligned} \frac{dt}{t} &= \frac{-(2p - 1)}{pr y_t + p(1 - r) - (2p - 1)} \frac{dy_t}{y_t} \\ &= \frac{-(2p - 1)}{pr y_t + pr(f_c - 1)} \frac{dy_t}{y_t} \\ &= -\frac{2p - 1}{pr} \left\{ \frac{1}{y_t + (f_c - 1)} \right\} \frac{dy_t}{y_t} \\ &= -\frac{2p - 1}{pr(f_c - 1)} \left\{ \frac{f_c - 1}{y_t + (f_c - 1)} \right\} \frac{dy_t}{y_t} \\ &= -\frac{2p - 1}{pr(f_c - 1)} \left\{ 1 - \frac{y_t}{y_t + (f_c - 1)} \right\} \frac{dy_t}{y_t} \\ &= -\frac{2p - 1}{pr(f_c - 1)} \left\{ \frac{1}{y_t} - \frac{1}{y_t + (f_c - 1)} \right\} dy_t. \end{aligned}$$

Hence, for an appropriate constant c , we have

$$c + \log t = \frac{2p - 1}{pr(f_c - 1)} [\log(y_t + (f_c - 1)) - \log y_t] = \frac{2p - 1}{pr(f_c - 1)} \log\left(1 + \frac{1 - r}{y_t}\right),$$

and so

$$1 + \frac{f_c - 1}{y_t} = \exp \left\{ \{c + \log t\} \frac{pr(f_c - 1)}{2p - 1} \right\} = Ct^\gamma,$$

where $\gamma = \frac{pr(f_c - 1)}{2p - 1} = \frac{1 - p - pr}{2p - 1}$ and $C = e^{c\gamma}$. Thus

$$y_t = \frac{f_c - 1}{Ct^\gamma - 1} \sim C't^{-\gamma}, \quad t \rightarrow \infty.$$

Moreover, the number of sites is approximately

$$rpty_t \sim C'r(1 - r)pt^{1-\gamma}.$$

Remark 2 For $f_c > 1$ we have $\gamma = \gamma(p, r) > 0$, and $\gamma(p, r)$ is a decreasing function of p . Also

- (i) when $p = 1 - p$, i.e., $p = \frac{1}{2}$, then $\gamma = \infty$; this corresponds to the case when the process dies out repeatedly,
- (ii) when $pr = 1 - p$, i.e., $f_c = 1$, then $\gamma = 0$; this corresponds to the case when the number of sites surviving is of order $o(t)$.
- (iii) when $p = \frac{2}{3+r} \in (\frac{1}{2}, \frac{1}{1+r})$, then $\gamma = 1$; this corresponds to the case when there are only a bounded number of sites surviving.

From the above, we see that there are three critical values

$$p_c^{(0)} := \frac{1}{2} < p_c^{(1)} := \frac{2}{3+r} < p_c^{(2)} := \frac{1}{r+1} < 1$$

and four phases:

1. For $p \in (p_c^{(2)}, 1)$, $\gamma \in (-r, 0)$ and individuals exist in the interval $(f_c, 1]$.
2. For $p \in (p_c^{(1)}, p_c^{(2)})$, $\gamma \in (0, 1)$ and the number of sites are increasing with the order $t^{1-\gamma}$ and the average number of particles is of order t^γ .
3. For $p \in (p_c^{(0)}, p_c^{(1)}]$, $\gamma \in (1, \infty)$, that is, $1 - \gamma$ is negative, and the number of sites is finite, with the average number of individuals being of order t .
4. For $p \in (0, p_c^{(0)}]$ the process dies out infinitely often.

7 Simulation

We conclude the paper with some simulations. The R code is given in the appendix. Here we have taken $p = 3/4$, $r = 1/2$, so that $f_c = 2/3$. The simulation has been conducted with $n = 100,000$.

Figure 1 presents the size of the population in \log_2 scale at each surviving site. The plot above the red line indicates the sites where the population size is 2^6 or

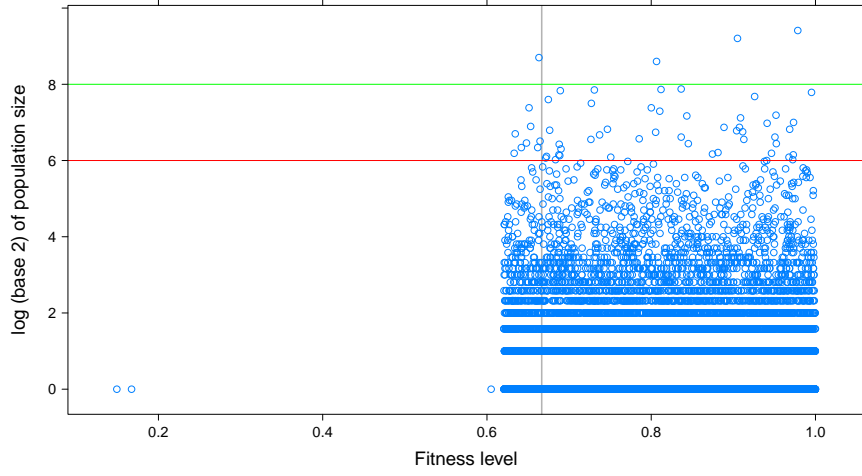


Figure 1: Population (in \log_2 scale) at various fitness levels.

more, while the plot above the green line indicates the sites where the population size is 2^8 or more.

In Figure 2 the x -axis gives the population size, while the y -axis presents the proportion of sites with the given population size. The blue line is the theoretical value as obtained from Theorem 3 and the vertical bars are the observed values.

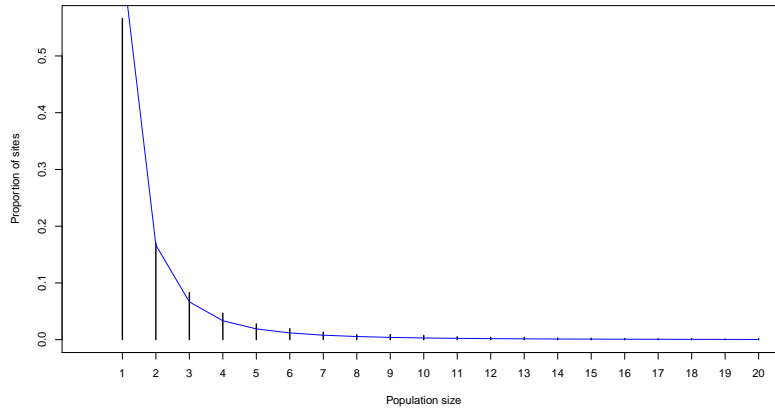


Figure 2: Theoretical and observed proportion of sites with respect to population size.

8 Acknowledgements

The authors are grateful to Professor Deepayan Sarkar who wrote the R code and performed the simulation. Rahul Roy acknowledges the grant MTR/2017/000141 from DST which supported this research and also the hospitality of Keio University where much of the work was done. Hideki Tanemura's research is supported in part by Grant-in-Aid for Scientific Research (S), No.16H06338; Grant-in-Aid for Scientific Research (B), No.19H01793 from Japan Society for the Promotion of Science.

References

- [1] Barabási, A. and Albert, R. : Emergence of scaling in random networks, *Science*, **286**, (1999), 509–512.
- [2] Benaïm, M., Schreiber, S. J. and Tarrès : Generalized urn models of evolutionary processes, **14**, (2004), 1455–1478
- [3] Ben-Ari, I. and Schinazi, R. B. : A stochastic model for the evolution of a quasispecies, *Journal of Statistical Physics*, **162**, (2016), 415–425.
- [4] Guiol, H., Machado, A. and Schinazi, R. B. : A stochastic model of evolution, *Markov processes and Related Fields* **17**, (2011), 253–258.
- [5] van der Hofstad, R. : *Random graphs and complex networks, Vol. 1*, (2017), Cambridge University Press, New York.
- [6] Liggett, T. M. and Schinazi, R. B. : A stochastic model for phylogenetic trees, *J. Appl. Probab.*, **46**, (2009), 601–607.
- [7] Schreiber, S. J. : Urn models, replicator processes, and genetic drift, *SIAM J. Appl. Math.*, **61** (2001), 2148–2167.

Appendix: The R code for the simulation

```
library(lattice)
library(latticeExtra)

createState <- function(MAX_POP = 10000L, p = 3/4, r = 1/2)
{
  n <- integer(MAX_POP) # size of each sub-population
  f <- numeric(MAX_POP) # fitness of each sub-population
  tob <- integer(MAX_POP) # time at which this population first appeared
```

```

n[1] <- 1L
f[1] <- 0
npop <- 1L
ndead <- 0L
t <- 0L
environment()v
}
## Make sure to keep normalized by ordering f from low to high
updateState <- function(S)
{
  p <- S$p
  r <- S$r
  S$t[] <- S$t + 1L # increment process lifetime counter
  u <- runif(1) # to decide which branch
  f <- runif(1) # new fitness value if needed
  if (u < 1-p) # kill particle with lowest fitness
  {
    if (S$n[1] > 0L) S$n[1] <- S$n[1] - 1L
    if (S$npop > 0 && S$n[1] == 0L) { # a population has just died out
      S$ndead[] <- S$ndead + 1L
      S$n[1:S$npop] <- S$n[2:(S$npop+1)]
      S$f[1:S$npop] <- S$f[2:(S$npop+1)]
      S$tob[1:S$npop] <- S$tob[2:(S$npop+1)]
      S$npop[] <- S$npop - 1L
    }
  }
  else if (u < 1 - p + p * r || S$npop == 0) # create new sub-population\\
  {
    S$npop[] <- S$npop + 1L
    if (S$npop == S$MAX_POP)
      stop("exceeded maximum sub-populations allowed: ", S$MAX_POP)
    S$f[S$npop] <- f
    S$n[S$npop] <- 1L
    S$tob[S$npop] <- S$t
    i <- 1:S$npop
    ord <- order(S$f[i])
    S$n[i] <- (S$n[i])[ord]
    S$f[i] <- (S$f[i])[ord]
    S$tob[i] <- (S$tob[i])[ord]
  }
  else { # increment size of one population by 1
    i <- sample(S$npop, 1, prob = S$n[1:S$npop])
    S$n[i] <- S$n[i] + 1L
  }
}

```

```

    }
}

```

```

S <- createState(MAX_POP = 20000, p = 3/4, r = 1/2)
(f_c <- with(S, (1-p) / (p*r)))

```

```

for (i in 1:100000) updateState(S)

```

```

Sdf <- subset(as.data.frame(as.list(S)), n > 0, select = c(tob, f, n))
names(Sdf) <- c("time of birth", "fitness", "population size")
xyplot(log2('population size') ~ fitness, data = Sdf, cex = 0.7,
       ylab = "log (base 2) of population size", xlab = "Fitness level",
       abline = list(v = f_c, col = "grey70", lwd = 2)) + layer(panel.abline(h = c(6, 8)

```

```

pk <- function(k, r) 1 / (1-r) * beta((2-r) / (1-r), k)
plot(prop.table(table( Sdf[["population size"]] )),
     xlim = c(0, 20), xlab = "Population size", ylab = "Proportion of sites")
lines(1:20, pk(1:20, r = S$r), col = "blue")

```

Rahul Roy
 Theoretical Statistics and Mathematics Unit
 Indian Statistical Institute
 7 SJS Sansanwal Marg
 New Delhi 110016, India.

Hideki Tanemura
 Department of Mathematics
 Keio University
 Hiyoshi, Kohoku-ku
 Yokohama 2238522, Japan