A stochastic model for the evolution of species with random fitness

Daniela Bertacchi*   Jüri Lember†   Fabio Zucca‡

Abstract

We generalize the evolution model introduced by Guiol, Machado and Schinazi (2010). In our model at odd times a random number X of species is created. Each species is endowed with a random fitness with arbitrary distribution on $[0, 1]$. At even times a random number Y of species is removed, killing the species with lower fitness. We show that there is a critical fitness f_c below which the number of species hits zero i.o. and above of which this number goes to infinity. We prove uniform convergence for the fitness distribution of surviving species and describe the phenomena which could not be observed in previous works with uniformly distributed fitness.

Keywords: generalized GMS model; birth and death process; survival; fitness; queuing process; limit distribution.
AMS MSC 2010: Primary 60J20, Secondary 60J80; 60J15.
Submitted to ECP on April 19, 2018, final version accepted on November 7, 2018.

1 Introduction

During the history of our planet, species have emerged and have become extinct, some have lasted a relatively brief period, others are still present in a more or less unchanged form after millions of years. It is widely accepted that the driving engine of evolution is natural selection or "survival of the fittest". It is therefore interesting to provide mathematical models for the evolution of species.

Guiol, Machado and Schinazi [7] proposed a model where creation and deletion of species is driven by chance in the sense that at each step with probability p one new species is created and its fitness is chosen uniformly in $[0, 1]$, while with probability $1 - p$ the least fit species (if there are species alive at that time) is removed. One motivation for the study of this model is that its long-term behaviour is similar to the one which simulations show for the Bak-Sneppen model: there is a critical value for the fitness and species with smaller fitness disappear, while species with a larger fitness persist indefinitely. Bak and Sneppen [1] modelled a simple ecosystem where the population size is constant and at each step not only the least fit is removed, but also its neighbours are replaced by new species (proximity may be seen as representing ecological links

---
*Dipartimento di Matematica e Applicazioni, Università di Milano–Bicocca, via Cozzi 53, 20125 Milano, Italy.
 E-mail: daniela.bertacchi@unimib.it
†Institute of Mathematics and Statistics, University of Tartu, J. Liiv 2, 50409 Tartu, Estonia.
 E-mail: juri.lember@ut.ee
‡Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
 E-mail: fabio.zucca@polimi.it

between species). It has proven difficult to obtain rigorous results for this model (see for instance [9]) and this motivates the search for similar, more tractable models.

Several papers have studied the GMS model: [3] gives a law of the iterated logarithm and a central limit theorem for number of species with supercritical fitness which go extinct (this number is negligible with respect to n); [6] studies the maximal fitness ever appeared in the subcritical case.

The model has been generalized in [10] and [2]: there is still a toss of a coin to decide for creation or deletion, but instead of adding/removing one species at a time, increments are arbitrary random variables. Even with these assumptions, the same cut-off phenomenon of [7] appears.

In the original GMS, the lengths of subsequent births and deaths are geometrically distributed random variables (with parameters which sum up to 1) and in [2, 10] they are geometrical convolutions of certain laws (where the parameters of these geometrically distributed number of convolutions, again, sum up to 1). In our model we group all subsequent creations and deletions: the length of subsequent creations $\{X_n\}_{n\in\mathbb{N}}$ and the length of subsequent annihilations $\{Y_n\}_{n\in\mathbb{N}}$ are such that $\{(X_n, Y_n)\}_{n\in\mathbb{N}}$ is an i.i.d. sequence with arbitrary distribution. Whence our results apply to the models in [2, 7, 10] (see Section 2.1). Besides, in the older papers the fitness is assigned uniformly while we use a general distribution μ. If μ has atoms, a new phenomenon appears: there might be a fitness which acts as a barrier eventually protecting all species with higher fitness (see Corollary 2.4 and the subsequent discussion for details).

Here is the outline of the paper. In Section 2 we give the formal construction of the process and the necessary definitions. We state our main result, Theorem 2.2, which describe the asymptotic expression of the proportion of species in a generic (Borel) range of fitness. The asymptotic behavior of a single fitness is described by Theorem 2.3. Corollary 2.4 and the subsequent discussion gives some details on the number of species which are killed. Section 2.1 is devoted to a detailed comparison with previous works; we explain why our work is a generalization of the previous models and which new phenomena arise. At the end of the section we make a conjecture about a generalization of our model with a different killing strategy (see Remark 2.5).

In Section 3 we study an example of a Markov model which cannot be treated by using previously known results (see Section 3.1). We also give a counterexample to be compared with Theorem 2.2(2).

All the proofs are in Section 4 which contains a couple of results which are worth mentioning: a *Law of Large Numbers* (Proposition 4.2) and Proposition 4.1 which identifies the set of fitness which become empty i.o. (and the total amount of time they are empty).

2   The process and its asymptotic behaviour

We start by giving a formal description of the process.

Let $\{X_n, Y_n, f_{n,i}\}_{n,i\in\mathbb{N}}$ be a family of nonnegative random variables and, for all $n \in \mathbb{N}$, denote by f_n the sequence $\{f_{n,i}\}_{i\in\mathbb{N}}$. Suppose that

1. for every $n \in \mathbb{N}$, (X_n, Y_n) and f_n are independent,

2. $\{(X_n, Y_n, f_n)\}_{n\in\mathbb{N}}$ are i.i.d.

3. all $f_{n,i}$ are distributed according to a measure μ on $\mathbb{R}$.

Roughly speaking, X_n counts the new species, Y_n counts the deaths and $f_{n,i}$ the fitness of a newly created species. In order to avoid trivial cases we suppose that $\mathbb{E}[X_k]$ and $\mathbb{E}[Y_k]$ are both in $(0, +\infty]$; moreover we assume that at least one of these two expected values

is finite. Note that in this case $\{X_n\}_{n\in\mathbb{N}}$, $\{Y_n\}_{n\in\mathbb{N}}$ and $\{X_n - Y_n\}_{n\in\mathbb{N}}$ are all i.i.d. families, but X_i and Y_i might be dependent. From now on, we will denote by (X, Y) a couple with the same law as (X_1, Y_1). For every fixed $n \in \mathbb{N}$ also $\{f_{n,i}\}_{i\in\mathbb{N}}$ might be dependent (for instance they can be generated by a Markov Chain or $f_{n,1} = f_{n,i}$ for all $i \in \mathbb{N}$).

We will assume that $\mu([0,1]) = 1$; there is no loss of generality, since any measure on $\mathbb{R}$ can be mapped to a measure supported in $[0,1]$. We denote its cumulative distribution function by $F = F_\mu$ and we define $F(f^-) := \lim_{a\to f^-} F(a)$.

Let Z_n be the number of species alive at time n. We start at time 0 with $Z_0 = 0$ (Z_0 could be a random variable with an arbitrary distribution on $\mathbb{N}$).

At time 1, X_1 species are generated and to each of them we assign a random fitness with law μ. More precisely the fitness of the i-th created species is $f_{1,i}$ for all $1 \le i \le X_1$. Thus $Z_1 = Z_0 + X_1$. The procedure is repeated at any odd time: $Z_{2k+1} = Z_{2k} + X_{k+1}$, meaning that X_{k+1} species are created and their fitness $f_{k+1,1}, \dots, f_{k+1,X_{k+1}}$ are assigned. For any set $A \subseteq [0,1]$ we denote by $Z_n(A)$ the total number of species alive at time n and with fitness in A. The fitness of a species does not change during its entire lifetime, and species may disappear only at even times.

At time $2k+2$, a number $Y_{k+1} \wedge Z_{2k+1}$ of species are removed and removal starts from the least fit. This means that $Z_{2k+2} = 0 \vee (Z_{2k+1} - Y_{k+1})$. Thus if $Y_{k+1} \ge Z_{2k+1}$ then $Z_{2k+2}(A) = 0$ for all $A \subseteq [0,1]$. Otherwise, let $x_+ := \min\{x \in [0,1] \colon Z_{2k+1}([0,x]) \ge Y_{k+1}\}$. All species with fitness smaller than x_+ are removed and $Z_{2k+2}(A) = 0$ for all $A \subseteq [0, x_+)$. A number $M_{k+1} := Y_{k+1} - Z_{2k+1}([0, x_+))$ of species is removed from the set of species with fitness equal to x_+: $Z_{2k+2}(\{x_+\}) = Z_{2k+1}(\{x_+\}) - M_{k+1}$ and $Z_{2k+2}(A) = Z_{2k+1}(A)$ for all $A \subseteq (x_+, 1]$.

Given a Borel set $A \subseteq [0,1]$ such that $\mu(A) > 0$, we define the number of species created in A as

$$\widetilde{X}_n(A) = \sum_{i=1}^{X_n} \mathbb{1}_A(f_{n,i}); \tag{2.1}$$

to avoid a cumbersome notation, henceforth we simply write $\widetilde{X}_n$ instead of $\widetilde{X}_n(A)$. By our assumptions, for any A, we have that $\{(\widetilde{X}_n, Y_n)\}_{n\in\mathbb{N}}$ are i.i.d. and $\mathbb{E}[\widetilde{X}_n] = \mu(A)\mathbb{E}[X_n]$ (where $a \cdot (+\infty) = +\infty$, if $a > 0$ and $0 \cdot \infty = 0$). Henceforth, an interval $I \subseteq [0,1]$ (either closed or not) such that $0 \in I$ is called a *left interval*. We note that, for a left interval I such that $\mu(I) > 0$, $\{Z_{2n}(I)\}_{n\in\mathbb{N}}$ is the queuing process (see [5, Ch.VI.9]) associated to the i.i.d. increments $\{\widetilde{X}_n - Y_n\}_{n\in\mathbb{N}}$ (see Section 4 for details).

We will often make use of the expected value $\mathbb{E}[\alpha X - Y]$ where $\alpha \in [0,1]$. If $\mathbb{E}[X] = +\infty > \mathbb{E}[Y]$ and $\alpha > 0$ then $\mathbb{E}[\alpha X - Y] := +\infty$; if $\mathbb{E}[Y] = +\infty > \mathbb{E}[X]$ then $\mathbb{E}[\alpha X - Y] := -\infty$ for all $\alpha \in [0,1]$.

We define the critical parameter:

$$f_c := \inf\{f \in \mathbb{R} \colon F(f) > \mathbb{E}[Y]/\mathbb{E}[X]\} \tag{2.2}$$

Note that when $\mathbb{E}[Y] \ge \mathbb{E}[X]$ then $f_c = +\infty$, otherwise f_c is the only solution of $F(f_c) \ge \mathbb{E}[Y]/\mathbb{E}[X] \ge F(f_c^-)$, where both inequalities turn into equalities if and only if $\mu(\{f_c\}) = 0$.

When $\mathbb{E}[Y] < \mathbb{E}[X] < +\infty$, we define the following probability measure (on Borel sets $A \subseteq [0,1]$) and its cumulative distribution function

$$\mathbb{P}_\infty(A) := \frac{\mu(A \cap (f_c, 1])\mathbb{E}[X] + \mathbb{1}_A(f_c)\mathbb{E}[\mu([0, f_c))X - Y]}{\mathbb{E}[X - Y]},$$

$$F_\infty(f) := \begin{cases} 0 & f < f_c, \\ \frac{\mathbb{E}[F(f)X - Y]}{\mathbb{E}[X - Y]} & f \ge f_c. \end{cases} \tag{2.3}$$

Definition 2.1. *Let $A \subseteq [0,1]$. We say that*

(i) *there is extinction in A if and only if $Z_n(A) = 0$ infinitely often a.s.;*

(ii) *there is survival in A if and only if $Z_n(A)$ is eventually positive as $n \to +\infty$ a.s.;*

(iii) *there is persistence in A if and only if for all even $n \in \mathbb{N}$ we have $\mathbb{P}(Z_m(A) > 0, \forall m \geq n | Z_n(A) > 0) > 0$.*

When $A = \{f\}$ we speak of extinction, survival and persistence of the fitness f.

As a consequence of Theorem 2.2, for every Borel set A, either there is extinction in A or there is survival (and $Z_n(A) \to +\infty$ almost surely). Moreover, persistence can be proved also for sets A such that $\mathbb{P}(Z_n(A) > 0) = 0$ (such as some singletons for instance); clearly if $\mathbb{P}(Z_n(A) > 0) > 0$ for some even n then persistence implies survival.

Theorem 2.2 (**Limit Distribution Theorem**).

(1) *For all sets $A \subseteq [0,1]$ such that $\mu(A \setminus [0, f_c)) = 0$, there is extinction in A and $Z_n(A)/n \overset{n \to +\infty}{\longrightarrow} 0$ uniformly with respect to A almost surely. If $F(f_c) = \mathbb{E}[Y]/\mathbb{E}[X]$ then the same holds for all $A \subseteq [0,1]$ such that $\mu(A \setminus [0, f_c]) = 0$.*

(2) *If, for every n, $\{f_{n,i}\}_{i \in \mathbb{N}}$ are i.i.d and $\mathbb{E}[X] = +\infty > \mathbb{E}[Y]$ then we have that $Z_n/n \overset{n \to +\infty}{\longrightarrow} \infty$ and $Z_n(A)/Z_n \overset{n \to +\infty}{\longrightarrow} \mu(A)$ a.s. (for Borel sets A such that $\mu(A) > 0$).*

(3) *If $\mathbb{E}[X - Y] \in (0, +\infty)$ then $Z_n/n \overset{n \to +\infty}{\longrightarrow} \mathbb{E}[X - Y]/2$ a.s. and*

$$\mathbb{P}\left(\frac{Z_n(A)}{Z_n} \overset{n \to +\infty}{\longrightarrow} \mathbb{P}_\infty(A), \text{ for every Borel set } A \subseteq [0,1] \right) = 1. \tag{2.4}$$

Moreover

$$\sup_{f \in [0,1]} \left| \frac{Z_n([0,f])}{Z_n} - F_\infty(f) \right| \to 0, \qquad \text{as } n \to +\infty, \ a.s.$$

It is worth noting that, as a consequence of Theorem 2.2(1), whenever $\mathbb{E}[\mu(I)X - Y] \in [-\infty, 0]$ for some left interval I, then $Z_{2n}(I) = 0$ infinitely often a.s.; nevertheless $Z_{2n}(I)$ has a non-trivial limit in law (see Proposition 4.1(3) for details). This implies that when $\mathbb{E}[X] < +\infty$ and $\mathbb{E}[X] \leq \mathbb{E}[Y] \leq +\infty$ then all fitness go extinct.

The example given in Section 3.2 shows that, if $\{f_{n,i}\}_{i \in \mathbb{N}}$ are just dependent, then the conclusion in Theorem 2.2(2) does not hold in general. Nevertheless, independence is not necessary as explained at the end of Section 3.2.

The following theorem describes the long-term behaviour of a fixed fitness. Note that all $f > f_c$ belong to case (1), while all $f < f_c$ belong to (2). If $f = f_c$, then case (2) applies if and only if $F(f_c) = \mathbb{E}[Y]/\mathbb{E}[X]$.

Theorem 2.3 (**Extinction, survival and persistence**). *Let $f \in [0,1]$.*

(1) *If $\mathbb{E}[F(f)X - Y] \in (0, +\infty]$ then there is persistence and survival in $[0, f]$ and the fitness f is persistent. Moreover, $\lim_{n \to \infty} Z_n([0, f]) = \infty$ a.s. and, if $\mu(\{f\}) > 0$ then $\lim_{n \to \infty} Z_n(\{f\}) = \infty$ almost surely.*

(2) *If $\mathbb{E}[F(f)X - Y] \in [-\infty, 0]$ then there is extinction in $[0, f]$.*

It could be shown that if $\mathbb{P}(Y = 0) > 0$ then the persistence in Theorem 2.3 holds in a stronger way, namely by removing the request of "even n" in Definition 2.1(iii). Indeed, in this case, given any A, $\mathbb{P}(Z_{2i+2}(A) > 0 | Z_{2i+1}(A)) \geq \mathbb{P}(Y = 0) > 0$ and Theorem 2.3 applies.

Denote by $K_n(A)$ the number of species killed in A up to time n and by $\tau_n(A)$ the total number of epochs that there are no species in A up to time n. From Theorem 2.2, if $\mathbb{E}[X - Y] > 0$ then, as $n \to +\infty$,

$$\frac{K_n(A)}{n} = \frac{\sum_{i=1}^{\lfloor (n+1)/2 \rfloor} \widetilde{X}_i - Z_n(A)}{n} \sim \frac{1}{2}\mu(A \cap [0, f_c])\mathbb{E}[X] - \frac{1}{2}\mathbb{1}_A(f_c)\mathbb{E}[\mu([0, f_c])X - Y] \qquad a.s. \tag{2.5}$$

where $\widetilde{X}_n$ is the number of species created in the Borel set A (see equation (2.1)).

Corollary 2.4. *If $\mathbb{E}[X - Y] > 0$ then*

(1) $\lim\limits_{n \to +\infty} K_n((f_c, 1])/n = 0$ *a.s.;*

(2) $K_n([f_c, 1]) \to +\infty$ *a.s.;*

(3) *If $F(f_c) > \mathbb{E}[Y]/\mathbb{E}[X]$ then $\sup\limits_{n \in \mathbb{N}} K_n((f_c, 1]) < +\infty$ a.s., otherwise $\lim\limits_{n \to +\infty} K_n((f_c, 1]) = +\infty$ a.s.;*

(4) *If $F(f_c^-) < \mathbb{E}[Y]/\mathbb{E}[X]$ then $\lim\limits_{n \to +\infty} K_n([f_c, 1])/n > 0$ a.s., otherwise $\lim\limits_{n \to +\infty} K_n([f_c, 1])/n = 0$ a.s.;*

(5) *If $f > f_c$ then $\sup\limits_{n \in \mathbb{N}} K_n([f, 1]) < +\infty$ almost surely.*

Here is a more explicit description. First of all, by (5) a.s. there are no more species killed in $[f, 1]$ eventually as $n \to +\infty$ but by (2) the number of species killed in $[f_c, 1]$ diverges almost surely.

If $\mu(\{f_c\}) = 0$ then $F(f_c) = \mathbb{E}[Y]/\mathbb{E}[X] = F(f_c^-)$, so that by (4) $K_n([f_c, 1])/n$ goes to zero a.s. and $\tau_n([0, f_c])/n \to 0$ almost surely as $n \to +\infty$ (see Proposition 4.1(2)).

If $\mu(\{f_c\}) > 0$ then we have the following possibilities:

- $F(f_c) > \mathbb{E}[Y]/\mathbb{E}[X] = F(f_c^-)$ then by (4) $K_n([f_c, 1])/n$ goes to zero almost surely. Moreover, by Theorem 2.3(1) we have $Z_n(\{f_c\}) \to +\infty$ almost surely as $n \to +\infty$, implying that the species killed in $[f_c, 1]$ eventually will have fitness f_c almost surely. Even though the number of species of fitness f_c which are killed diverges, by equation (4.2) $Z_n(\{f_c\})/n$ converges to $(F(f_c) - \mathbb{E}[Y]/\mathbb{E}[X]) \cdot \mathbb{E}[X]/2 > 0$. Also, $\tau_n([0, f_c))/n$ goes to 0 (see again Proposition 4.1(2)).

- $F(f_c) > \mathbb{E}[Y]/\mathbb{E}[X] > F(f_c^-)$ then, just as before, a.s. the species killed in $[f_c, 1]$ eventually will have fitness f_c and the fraction of species alive with fitness f_c converges to the same positive limit. This time $K_n(\{f_c\})/n$ has a positive limit: $-\mathbb{E}[F(f_c^-)X - Y]/2$ and $\tau_n([0, f_c))/n$ converges to a positive limit almost surely as $n \to +\infty$ (see Proposition 4.1(3)).

- $F(f_c) = \mathbb{E}[Y]/\mathbb{E}[X] > F(f_c^-)$ then, by Theorem 2.2, every species with fitness f_c is eventually killed a.s. and $K_n([f_c, 1])/n$ converges to $-\mathbb{E}[F(f_c^-)X - Y]/2 > 0$. But $K_n((f_c, 1])/n$ tends to 0, a.s., thus outside a negligible proportion, the killed species all have fitness f_c, whence $K_n(\{f_c\})/n$ has the same positive limit as before. Finally, $\tau_n([0, f_c])/n \to 0$ almost surely as $n \to +\infty$ (see Proposition 4.1(2)).

2.1 Comparison with previous works

Our process extends those appeared in [2, 7, 10]. Aside from our general choice for the fitness law, the birth-and-death mechanism that we study is more general than those adopted in these papers.

One way to see the original GMS (see [7]) as a particular case of our process is to consider the variables (X_n, Y_n) taking only the values $(1, 0)$ and $(0, 1)$ with probability p and $1 - p$ respectively.

A different way to describe the GMS by means of our framework is by observing that the random sequences of consecutive births X_k and consecutive deaths Y_k have right-shifted Geometric distribution with parameter $1-p$ and p respectively.

In general, consider a process $\{\overline{Z}_n\}_{n\in\mathbb{N}}$ where at each step either a species is created (along with its fitness) or the least-fit species, if any, is removed. Denote by $X_1 > 0$ the length of the first stretch of "creations", followed by a stretch of "annihilations" of length $Y_1 > 0$, then another stretch of "creations" of length X_2 followed by a stretch of Y_2 "annihilations" and so on. Suppose that $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are two i.i.d. sequences. It is clear that there is a connection between our process and this one, namely for every set A, $Z_n(A) = \overline{Z}_{N_n}(A)$ where $N_n = \sum_{i=1}^{\lfloor(n+1)/2\rfloor} X_i + \sum_{i=1}^{\lfloor n/2\rfloor} Y_i$.

In particular if n is even and $k \in (N_n, N_{n+1})$, then $\overline{Z}_k(A)$ is nondecreasing, while if $k \in (N_{n+1}, N_{n+2})$, then $\overline{Z}_k(A)$ is nonincreasing. Proposition 4.2 shows that for every left interval I

$$\frac{Z_n(I)}{n} = \frac{\overline{Z}_{N_n}(I)}{n} \to \frac{\mathbb{E}[\mu(I)X - Y] \vee 0}{2}, \quad \text{a.s.}$$

When $\mathbb{E}[X + Y] < \infty$, then the monotonicity of $\overline{Z}_k$ between N_n and N_{n+1}, implies

$$\frac{\overline{Z}_n(I)}{n} \to \frac{\mathbb{E}[\mu(I)X - Y]}{2} \frac{2}{\mathbb{E}[X + Y]} = \frac{\mathbb{E}[\mu(I)X - Y]}{\mathbb{E}[X + Y]}, \quad \text{a.s.}$$

Therefore, the long-term behaviour of $\{\overline{Z}_n(I)\}_{n\in\mathbb{N}}$ can be derived simply by studying $\{Z_n(I)\}_{n\in\mathbb{N}}$.

Our work can also be considered as a generalization of [2] and [10] whose models are essentially equivalent. Indeed, in [2], a single family of $\mathbb{Z}$-valued variables $\{U_n\}_{n\in\mathbb{N}}$ is considered. In this process, $U_n > 0$ means that U_n species are created, while $U_n < 0$ means that $-U_n$ species are killed. In this case the laws of length of a "creation" stretch X_i and "annihilation" stretch Y_i are necessarily geometric random convolutions of the law of U_n conditioned on $\{U_n > 0\}$ and $\{U_n < 0\}$ respectively. Moreover, the sum of the parameters of these geometric convolutions must be $1 - \mathbb{P}(U_1 = 0)$. Therefore, a model constructed from the variables $\{U_n\}_{n\in\mathbb{N}}$ can be considered as a particular case of our model: take for instance $X_n := U_n \mathbb{1}_{\{U_n > 0\}}$, $Y_n := -U_n \mathbb{1}_{\{U_n < 0\}}$ and consider the process $\{Z_{2n}(A)\}_{n\in\mathbb{N}}$. In Section 3.1 we consider a particular case of our process which cannot be obtained with a single family of variables describing simultaneously creations and annihilations.

Observe that in Theorem 2.2 we used Z_n as a normalizing factor for $Z_n(A)$ but there are two other natural choices: n (to compare with [2, 10]) and N_n (to compare with [3, 7]).

If $\mathbb{E}[X_i + Y_i] < +\infty$ then, by the Strong Law of Large Numbers (SLLN), $N_n \sim n\mathbb{E}[X + Y]/2$ almost surely as $n \to +\infty$. If, in addition, $\mathbb{E}[X - Y] \in (0, +\infty)$ then by Proposition 4.2 we have

$$Z_n \sim n\mathbb{E}[X - Y]/2 \sim N_n \frac{\mathbb{E}[X - Y]}{\mathbb{E}[X + Y]} \qquad a.s.$$

as $n \to +\infty$. Hence Theorem 2.2(3) can be equivalently written in terms of the timescale n or N_n (in this last case we obtain a generalization of Proposition 4.2(1) to Borel sets). If $\mathbb{E}[X] = +\infty > \mathbb{E}[Y]$ then $Z_n \sim N_n$ almost surely as $n \to +\infty$. Indeed one can use the same kind of arguments used in the proof of Theorem 2.2(2), to prove that Z_n and $\sum_{i=1}^{\lfloor(n+1)/2\rfloor} X_i$ are asymptotic and the remaining terms are negligible. Roughly speaking, changing timescale turns out to be just a linear rescaling.

We note that for the GMS model and its generalizations, with $\mu \sim \mathcal{U}([0, 1])$ (where $\mathcal{U}(I)$ is the uniform distribution on I), the fraction of surviving species in any $I \subseteq [f_c, 1]$ is proportional to $\mu(I)$. This is still true in our case when $I \subseteq (f_c, 1]$, but it does not hold for instance if $I = [f_c, b]$ and $F(f_c^-) < \mathbb{E}[Y]/\mathbb{E}[X]$. Moreover if $\mu \sim \mathcal{U}([0, 1])$ then

$K_n([f_c, 1])/n \to 0$ (the exact rate of convergence for the GMS is studied in [3]), while again this needs not to be true if f_c is an atom for μ.

Remark 2.5. Consider our model with a different killing strategy: given Y_n species to be killed, for each one, independently, we choose to remove either the least fit species (with probability r) or one of the species chosen uniformly at random (with probability $1 - r$). We conjecture that $f_c = \inf\{f \in \mathbb{R}: \mathbb{E}[F(f)X - rY] > 0\}$ and Theorem 2.2 still holds.

3 Examples and counterexamples

3.1 The Markov case

Let the birth-death process be a Markov chain with transition matrix

$$\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

starting from a birth. Thus the probability of a birth after the birth $P_{++} = p$, the probability of death after the birth is $P_{+-} = 1 - p$ and so on. This can be seen as a particular case of our process where X has a geometric law $\mathcal{G}(1-p)$ while Y has a geometric law $\mathcal{G}(1-q)$. We assume that $p > q$; clearly $\mathbb{E}[Y]/\mathbb{E}[X] = (1-p)/(1-q)$.

We choose $\mu := \alpha\delta_{1/2} + (1-\alpha)\nu$ (where $\nu \sim \mathcal{U}([0, 1])$); thus the cumulative distribution function is $F(f) = (1 - \alpha)f + \alpha\mathbb{1}_{[1/2, 1]}(f)$.

Now $\mathbb{E}[\mu(I)X - Y] = \mu(I)/(1-p) - 1/(1-q)$ and, according to equation (2.3), $Z_n([0, f])/Z_n \to \mathbb{1}_{[f_c, 1]}(f)(F(f)\frac{1-q}{p-q} - \frac{1-p}{p-q})$ as $n \to +\infty$, where f_c is given by equation (2.2) and it is the unique solution in $[0, 1]$ of $F(f_c) \geq \mathbb{E}[Y]/\mathbb{E}[X] = (1-p)/(1-q) \geq F(f_c^-)$.

To avoid useless complications, we discuss just the "fair coin" case $\alpha = 1/2$. In this case we have

$$f_c = \begin{cases} 2(1-p)/(1-q) & (q+3)/4 < p \leq 1 \\ 1/2 & (1+3q)/4 \leq p \leq (3+q)/4 \\ (1+q-2p)/(1-q) & q < p < (1+3q)/4. \end{cases}$$

There are five typical situations represented by the five examples in the following table.

(p, q)	f_c	$F(f_c)$	$F(f_c^-)$	$\mathbb{E}[F(f_c)X - Y]$	$\mathbb{E}[F(f_c^-)X - Y]$	$\lim_{n\to+\infty} Z_n([0, f])/Z_n$ c.d.f.	Law
$(2/9, 1/9)$	$3/4$	$7/8$	$7/8$	0	0	$(4f - 3)\mathbb{1}_{[3/4, 1]}(f)$	$\mathcal{U}([3/4, 1])$
$(2/5, 1/5)$	$1/2$	$3/4$	$1/4$	0	$-5/6$	$(2f - 1)\mathbb{1}_{[1/2, 1]}(f)$	$\mathcal{U}([1/2, 1])$
$(3/4, 1/2)$	$1/2$	$3/4$	$1/4$	$3/4$	$-3/4$	$f\mathbb{1}_{[1/2, 1]}(f)$	$\frac{1}{2}\delta_{1/2} + \frac{1}{2}\mathcal{U}([1/2, 1])$
$(5/6, 1/3)$	$1/2$	$3/4$	$1/4$	3	0	$\frac{2f+1}{3}\mathbb{1}_{[1/2, 1]}$	$\frac{2}{3}\delta_{1/2} + \frac{1}{3}\mathcal{U}([1/2, 1])$
$(9/10, 1/5)$	$1/4$	$1/8$	$1/8$	0	0	$\frac{4f-1}{7}\mathbb{1}_{[1/4, 1]}(f) + \frac{4}{7}\mathbb{1}_{[1/2, 1]}(f)$	$\frac{4}{7}\delta_{1/2} + \frac{3}{7}\mathcal{U}([1/4, 1])$

3.2 Counterexample of Theorem 2.2(2) for dependent $\{f_{n,i}\}_{i\in\mathbb{N}}$

We define $Y_n := 1$ a.s. and $f_{n,i} := f_{n,1}$ for all $i > 1$ and $n \in \mathbb{N}$ (where $\{f_{n,1}\}_{n\in\mathbb{N}}$ is an i.i.d. sequence distributed according to μ). We construct the sequence $\{X_n\}_{n\in\mathbb{N}}$ as $X_n := g(H_n)$, for a suitable choice of an i.i.d. sequence $\{H_n\}_{n\in\mathbb{N}}$ and a function g.

Let $\{H_n\}_{n\in\mathbb{N}}$ be an i.i.d. sequence such that $\mathbb{P}(H_n = i) := 1/2^i$ for all $i \in \mathbb{N} \setminus \{0\}$. We define $n_i := i(i+1)/2$ for all $i \in \mathbb{N}$, hence $\mathbb{P}(H_n \leq n_{i+1}|H_n > n_i) = 1 - 1/2^{i+1}$ for all $i \in \mathbb{N}$.

Let us define $T_k := \min\{i: H_i > n_k\}$ for all $k \in \mathbb{N}$ (clearly $T_0 = 1$); note that $T_k \sim \mathcal{G}(1/2^{n_k})$. Since $\lim_{m\to+\infty} \mathbb{P}(T_k \leq m, H_{T_k} \leq n_{k+1}) = \mathbb{P}(H_{T_k} \leq n_{k+1}) = 1 - 1/2^{k+1}$ then there exists $\tau_k \in \mathbb{N}$ such that $\mathbb{P}(T_k \leq \tau_k, H_{T_k} \leq n_{k+1}) \geq 1 - 1/2^k$. The sequence $\{\tau_k\}_{k\geq 1}$ can be always constructed iteratively as a nondecreasing sequence. It is not

difficult to prove, by using the Borel-Cantelli Lemma, that the event $\Omega_0 := \bigcap_{k \in \mathbb{N}} \{H_{T_k} \leq n_{k+1}, T_k \leq \tau_k\}$ has positive probability.

We are now ready to define $g(i) := (k+1)! \prod_{j=0}^{k} \tau_j$ for all $i = n_k + 1, \ldots, n_{k+1}$ (for all $k \in \mathbb{N}$); $X_n := g(H_n)$ for all $n \in \mathbb{N}$. On Ω_0 we have

$$\frac{X_{T_k}}{\sum_{i=1}^{T_k} X_i} \geq \frac{X_{T_k}}{(T_k - 1)g(n_k) + X_{T_k}} \geq \frac{(k+1)! \prod_{j=0}^{k} \tau_j}{\tau_k \, k! \prod_{j=0}^{k-1} \tau_j + (k+1)! \prod_{j=0}^{k} \tau_j} \geq \frac{k+1}{k+2}.$$

Roughly speaking, this means that, on Ω_0, for every $\varepsilon > 0$ infinitely often the last generation represents at least a fraction $1 - \varepsilon$ of the entire population. Whence due to our choice of $\{f_{n,i}\}_{i \in \mathbb{N}}$, on Ω_0, for every Borel set $A \subseteq [0,1]$ such that $\mu(A) > 0$ and for every $\varepsilon > 0$, we have $Z_n(A)/Z_n \geq 1 - \varepsilon$ infinitely often.

It is worth noting, though, that independence is not necessary for the limit in Theorem 2.2(2) to hold. Indeed, consider a variable X taking just even, strictly positive values. Suppose that $f_{n,2i+1} = f_{n,2i+2}$ for all $i \geq 0$ but $\{f_{n,2i+1}\}_{i \in \mathbb{N}}$ are independent; roughly speaking, a random number of "coupled" species are generated and inside each couple the same random fitness is assigned. It is easy to show that $Z_n(A)/Z_n \to \mu(A)$ by the same argument used in the independent case (see the proof of Theorem 2.2(2)).

4 Proofs

We note that, for any fixed left interval I, $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ is a random walk on $\mathbb{N}$ (with increments depending on the position). More precisely it is the queuing process (see [5, Ch.VI.9]) associated to the increments $\widetilde{X}_n - Y_n$, as defined in equation (2.1). More precisely

$$Z_{2n+2}(I) - Z_{2n}(I) = \max(-Z_{2n}(I), \widetilde{X}_{n+1} - Y_{n+1}).$$

We denote by $\{S_n(I)\}_{n \in \mathbb{N}}$ the random walk with independent increments, where $S_n(I) := \sum_{i=1}^{n} (\widetilde{X}_i - Y_i)$. The drift of this random walk is $\mathbb{E}[\widetilde{X}_i - Y_i] = \mathbb{E}[\mu(I)X_i - Y_i]$ which is independent of i.

At time 0 we have $S_0 = Z_0(I) = 0$ and for all n

$$Z_{2n}(I) = S_n(I) - \min_{i \leq n} S_i(I) = \max_{i \leq n} \sum_{k=i+1}^{n} (\widetilde{X}_k - Y_k), \quad \forall n \in \mathbb{N}. \tag{4.1}$$

By the duality principle of random walks, using that $S_0(I) = 0$, the paths $(0, S_1(I), S_2(I), \ldots, S_n(I))$ and $(0, S_n(I) - S_{n-1}(I), S_n(I) - S_{n-2}(I), \ldots, S_n(I))$ have the same law (note that it follows from time-reversal of the increments, see also [4, Ch.III.8]). Thus, by equation (4.1), $Z_{2n}(I)$ and $\max_{k \leq n} S_k(I)$ have the same law. Since $S_0(I) = 0$ then $\min_{i \leq n} S_i(I) \leq 0$, hence $Z_{2n}(I) \geq S_n(I)$ for all $n \in \mathbb{N}$.

Define $d := \mathrm{GCD}(n \in \mathbb{Z} : \mathbb{P}(\widetilde{X} - Y = n) > 0)$; by elementary number theory it is easy to show that, since $\mathbb{E}[X]\mathbb{E}[Y] > 0$, when $\mu(I) > 0$ the random walk $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ (resp. $\{S_n(I)\}_{n \in \mathbb{N}}$) is irreducible on the set $\{dn : n \in \mathbb{N}\}$ (resp. $\{dn : n \in \mathbb{Z}\}$).

We start with the classification of the random walk $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$.

Proposition 4.1 (**Recurrence and transience**). *Let I be a left interval such that $\mathbb{P}(\widetilde{X} \neq Y) > 0$. Denote by τ_n the time spent at 0 by the random walk $\{Z_{2i}(I)\}_{i \in \mathbb{N}}$ up to time n. The random walk is*

(1) transient if and only if $\mathbb{E}[\mu(I)X - Y] \in (0, +\infty]$, in this case $\mathbb{P}(\sup_{n \in \mathbb{N}} \tau_n < +\infty) = 1$;

(2) null recurrent if and only if $\mathbb{E}[\mu(I)X - Y] = 0$, in this case $\mathbb{P}(\lim_{n \to +\infty} \tau_n/n = 0) = 1$;

(3) positive recurrent if and only if $\mathbb{E}[\mu(I)X - Y] \in [-\infty, 0)$, in this case $\mathbb{P}(\lim_{n \to +\infty} \tau_n / n > 0) = 1$. Moreover, as $n \to +\infty$

$$Z_{2n}(I) \overset{dist}{\to} S_\infty(I), \text{ a.s., where } S_\infty(I) := \sup_{n \geq 0} S_n(I) < \infty, \quad \text{a.s.}.$$

Note that the case where $\mathbb{P}(\widetilde{X} = Y) = 1$ is trivial, since it means that $\mu(I) = 1$ and $X = Y = c \in (0, +\infty)$ a.s.; thus, $Z_n(I)$ equals c when n is odd and 0 when n is even.

Proof. Recall the relation between the random walks $\{S_n(I)\}_{n \in \mathbb{N}}$ and $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ given by equation (4.1). In particular the return times to 0 of the second process are the weak descending ladder times of the first one, that is, the times n such that $S_n(I) \leq S_i(I)$ for all $i \leq n$. We denote by $\{T_i\}_{i \in \mathbb{N}}$ the sequence of intervals between two consecutive weak descending ladder times of $\{S_n(I)\}_{n \in \mathbb{N}}$ (that is, the times between two consecutive returns at 0 of $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$). Note that $\{T_i\}_{i \in \mathbb{N}}$ are i.i.d random variables and $T_1 = \min\{n \geq 1 : S_n \leq 0\}$.

(1) If $\mathbb{E}[\mu(I)X - Y] > 0$ (either finite or infinite) then, by the SLLN, $S_n(I) \to +\infty$ a.s., hence the same happens to the process $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ since $Z_{2n}(I) \geq S_n(I)$ for all $n \in \mathbb{N}$ (see equation (4.1) and the remark afterwards). This implies that $\inf_{n \geq 0} S_n(I) =: S_{-\infty} > -\infty$ a.s. and the Markov chain $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ is transient. As a consequence $\mathbb{P}(\sup_{n \in \mathbb{N}} \tau_n < +\infty) = 1$.

(2) When the distribution of $\widetilde{X}_n - Y_n$ is not degenerate (that is, it is not δ_0) then according to [5, Theorem 4, Ch.VI.10] $\{S_n(I)\}_{n \in \mathbb{N}}$ is a recurrent random walk on the set $\{dn : n \in \mathbb{Z}\}$. Since there are infinitely many reachable states on the left (as well as on the right) of the origin, we have $\mathbb{P}(T_1 < \infty) = 1$ so that $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ is recurrent. Moreover $\mathbb{E}[T_1] = +\infty$ (see [5, Theorem 2(i), Ch.XII.2]) and this implies the null recurrence of $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$. It is well known that, for a recurrent random walk $\mathbb{P}(\lim_{n \to +\infty} \tau_n / n = 1/\mathbb{E}[T_1]) = 1$ where, in this case, $1/\mathbb{E}[T_1] = 0$.

(3) We apply again the SLLN to $\{S_n(I)\}_{n \in \mathbb{N}}$ to deduce that $S_n(I) \to -\infty$ a.s., hence $\sup_{n \geq 0} S_n(I) =: S_\infty(I) < +\infty$ and $\mathbb{E}[T_1]$ is finite (see [5, Theorem 2(ii), Ch.XII.2]). Thus, $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ is positive recurrent. As before $\mathbb{P}(\lim_{n \to +\infty} \tau_n / n = 1/\mathbb{E}[T_1]) = 1$ where, in this case, $1/\mathbb{E}[T_1] > 0$.

It is clear that $\max_{i \leq n} S_n(I) \uparrow S_\infty(I)$ a.s. and the conclusion follows by equation (4.1) (see also [5, Ch.VI.9]). $\square$

The next proposition deals with the a.s. convergence of $Z_n(I)/n$ as $n \to \infty$.

Proposition 4.2 (**Law of large numbers**).

(1) For every interval $I \subseteq [0, 1]$,

$$Z_n(I)/n \to \frac{1}{2}\Big(\mu(I \cap (f_c, 1])\mathbb{E}[X] + \mathbb{1}_I(f_c)\mathbb{E}[\mu([0, f_c])X - Y]\Big), \quad \text{a.s.} \tag{4.2}$$

(2) If $\mathbb{E}[X - Y] \in [-\infty, 0]$ then, for all sets $A \subseteq [0, 1]$, $Z_n(A)/n \to 0$ almost surely as $n \to +\infty$.

(3) Let I be a left interval and $J \subseteq [0, 1]$ be such that $I \cap J = \emptyset$ and $\mu(J) > 0$. Suppose that $\mathbb{E}[\mu(I)X - Y] \in (0, +\infty]$. Then, a.s., $Z_n(J)$ is nondecreasing eventually as $n \to +\infty$ and $Z_n(J)/n \to \mu(J)\mathbb{E}[X]/2$.

Proof. (1) **Equation** (4.2) **for a left interval** I. In this case equation (4.2) becomes

$$\frac{Z_n(I)}{n} \to \frac{1}{2}\mathbb{E}[\mu(I)X - Y] \vee 0, \quad \text{a.s.} \tag{4.3}$$

Let $\Delta := \mathbb{E}[\mu(I)X - Y] \in [-\infty, +\infty]$; by the SLLN we have that **(a)** $S_n(I)/n \to \Delta$ a.s.. We separate two cases.

• $\Delta \in [-\infty, 0]$. Since $\liminf_n S_n(I) = -\infty$, then **(b)** for every n_0 there is a larger weak descending ladder time, i.e. $n \geq n_0$ such that $S_n(I) \leq S_k(I)$ for all $k \leq n$. Hence almost every trajectory satisfies both **(a)** and **(b)**; let us consider such a trajectory.

When $0 \geq \Delta > -\infty$ then for every $\varepsilon > 0$ there exists n_0 such that for every $n \geq n_0$ we have $|S_n(I)/n - \Delta| < \varepsilon/2$. Consider a weak descending ladder time $n_1 \geq n_0$; it is clear that, for every $n \geq n_1$ then $\min_{k \leq n} S_k(I) = S_{k_n}(I)$ for some k_n such that $n \geq k_n \geq n_1 \geq n_0$.

When $\Delta = 0$, then for every $n \geq n_1$ we have, by equation (4.1),

$$\frac{|Z_{2n}(I)|}{n} = \left| \frac{S_n(I)}{n} - \frac{\min_{k \leq n} S_k(I)}{n} \right| = \left| \frac{S_n(I)}{n} - \frac{S_{k_n}(I)}{n} \right| \leq \frac{|S_n(I)|}{n} + \frac{|S_{k_n}(I)|}{k_n} \cdot \frac{k_n}{n} < \varepsilon$$

since $n \geq n_0$ and $n \geq k_n \geq n_0$.
When $-\infty < \Delta < 0$, take $\varepsilon \in (0, -2\Delta)$. For every $n \geq n_1$ we have

$$\Delta + \varepsilon/2 \geq \frac{S_n(I)}{n} \geq \frac{S_{k_n}(I)}{n} \geq \frac{S_{k_n}(I)}{k_n} \geq \Delta - \varepsilon/2$$

since $S_{k_n}(I) \leq n(\Delta + \varepsilon) < 0$ (for all $n > 0$) and $n \geq k_n \geq n_0$. From the above chain of inequalities we obtain $|S_{k_n}(I)/n - \Delta| \leq \varepsilon/2$. Using again equation (4.1) we have

$$\frac{|Z_{2n}(I)|}{n} \leq \left| \frac{S_n(I)}{n} - \Delta \right| + \left| \frac{S_{k_n}(I)}{n} - \Delta \right| < \varepsilon.$$

If $\Delta = -\infty$, consider the process $\{\widehat{Z}_{2n}\}_{n \in \mathbb{N}}$ constructed by using $Y_k \wedge M$ instead of Y_k in such a way that $\mathbb{E}[\mu(I)X - Y \wedge M] \in (-\infty, 0)$. We have $0 \leq Z_{2n}(I)/n \leq \widehat{Z}_{2n}(I)/n \to 0$ almost surely as $n \to \infty$.

We are left to prove that

$$Z_{2n+1}(I)/n \to 0, \quad \text{a.s.} \tag{4.4}$$

Remember that, for all $\varepsilon > 0$, $\mathbb{E}[X] < \infty$ iff $\sum_{n \in \mathbb{N}} \mathbb{P}(X > \varepsilon n) < \infty$; thus, by the Borel-Cantelli's Lemma, $\mathbb{E}[X] < \infty$ implies $\mathbb{P}(\liminf_n \{X_n \leq \varepsilon n\}) = 1$. Thus

$$\frac{|Z_{2n+1}(I) - Z_n(I)|}{n} \leq \frac{X_{n+1}}{n} \to 0, \quad \text{a.s.} \tag{4.5}$$

so that from $Z_{2n}/n \to 0$, a.s., the convergence (4.4) follows.

• $\Delta \in (0, +\infty]$. By the SLLN, $S_n(I) \to +\infty$ a.s. and $\inf_{n \geq 0} S_n(I) =: S_{-\infty} > -\infty$, almost surely. By **(a)**, using equation (4.1), we have $Z_{2n}(I) = S_n(I) - S_{-\infty}$ eventually a.s., which implies $Z_{2n}(I)/n \to \mathbb{E}[\mu(I)X - Y]$, almost surely.

As before we are left to show

$$Z_{2n+1}(I)/n \to \mathbb{E}[\mu(I)X - Y], \quad \text{a.s.} \tag{4.6}$$

If $\mathbb{E}[X] < \infty$ then we use (4.5) to obtain (4.6).
If $\mathbb{E}[X] = \infty$, note that $Z_{2n+1}(I) \geq Z_{2n}(I)$, thus $Z_{2n+1}(I)/n \to +\infty$ almost surely.

Equation (4.2) **for a generic interval** I. Consider the the two left intervals $I_1 := \{x \in [0,1] : \exists y \in I, x \leq y\}$ and $I_2 := \{x \in [0,1] : x < y, \forall y \in I\}$. Clearly $I_2 \subseteq I_1$, $I_1 \setminus I_2 = I$, whence $Z_n(I) = Z_n(I_1) - Z_n(I_2)$. For I_1 and I_2, the convergence in equation (4.3) holds.

If $f_c \in I$, then $I_2 \subseteq [0, f_c)$ so that $E[\mu(I_2)X - Y] \leq 0$ and by the result for left intervals, $Z_n(I_1)/n \to 0$ almost surely. Therefore,

$$\lim_n \frac{Z_n(I)}{n} = \lim_n \frac{Z_n(I_1)}{n} = \frac{E[\mu(I_1)X - Y]}{2} = \frac{E[\mu(I \cup [0, f_c))X - Y]}{2}, \quad \text{a.s..}$$

Suppose $f_c \notin I$. If $I \subseteq [0, f_c)$, then $E[\mu(I_1)X - Y] \leq 0$ and $Z_n(I)/n \to 0$ almost surely. If, $I \subset (f_c, 1]$, then by equation (4.3) we have

$$\frac{Z_n(I)}{n} \to \frac{E[\mu(I)X - Y]}{2} = \frac{E[\mu(I \cap (f_c, 1])X - Y]}{2}, \quad \text{a.s.}$$

and the statement now follows. This concludes the proof of (1).

(2) In this case $f_c = +\infty$. Whence $Z_n/n \to 0$ a.s., thus the same holds for $Z_n(A)$ for every $A \subseteq [0, 1]$.

(3) The result follows from the fact that $\lim_{n \to +\infty} Z_n(I) = +\infty$ almost surely. Since $Z_n(I) \to +\infty$, a.s. then no species with fitness from $[0, 1] \setminus I$ are removed, eventually. By the SLLN, the number of births in J (up to time n) divided by n goes to its expectation almost surely as $n \to +\infty$ and this yields the claim. $\qquad\square$

Proof of Theorem 2.3. (1) Take $I = [0, f]$ and suppose that $Z_{2n_0}(I) > 0$. Since $Z_{2n+1}(I) \geq Z_{2n}(I)$ for all n, in order to check whether the process hits the origin or not, it is enough to consider the process $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$. By Proposition 4.1(1) $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ is transient and $\mathbb{P}(Z_{2n}(I) > 0, \ \forall n > n_0 | Z_{2n_0} = i) > 0$ (and this probability does not depend on n_0). Therefore we have persistence and survival. Moreover $Z_{2n}(I) \to \infty$ a.s., thus from $Z_{2n+1}(I) \geq Z_{2n}(I)$ we have $Z_n(I) \to \infty$ almost surely.

Suppose now that there are species with fitness f alive at time n_0. If $Z_n(I)$ is never empty for $n \geq n_0$, then the fitness f is never removed. Thus the persistence of f is equivalent to the persistence of I.

We now show that $Z_n(\{f\}) \to \infty$. If $Z_{2n}(\{f\}) > Z_{2n+2}(\{f\})$ then

$$Z_{2n}(I) \geq Z_{2n}(\{f\}) > Z_{2n+2}(\{f\}) = Z_{2n+2}(I).$$

For all $M > 0$ let $n(M)$ be a random integer such that $Z_{2n}(I) \geq M$ for all $n \geq n(M)$. For every $n \geq n(M)$ either $Z_{2n}(\{f\}) \leq Z_{2n+2}(\{f\})$ or $Z_{2n+2}(\{f\}) \geq M$. In particular, if $Z_{2n_1}(\{f\}) \geq M$ for some $n_1 > n(M)$, then it is true for all $n \geq n_1$. Consider the first (random) time $n_1 \geq n(M)$ when $Z_{2n_1}([0, f]) = 0$. If $n_1 = \infty$ then, from time $n(M)$ on, $Z_n(\{f\})$ is non decreasing and strictly increasing infinitely many often; indeed, species of fitness f are created infinitely many times a.s. (since $\mu(\{f\}) > 0$) and these species will be never removed after time $n(M)$ (since $Z_{2n}([0, f]) > 0$ for every $n \geq n(M)$). If, on the other hand, $n_1 < \infty$ then $Z_{2n}(\{f\}) \geq M$ for all $n \geq n_1$ and the result follows.

(2) By Proposition 4.1, applied to $I = [0, f]$, the process $\{Z_{2n}(I)\}_{n \in \mathbb{N}}$ is recurrent and so $Z_n(I) = 0$ infinitely often, almost surely. $\qquad\square$

Proof of Corollary 2.4. The statement (1) follows from the equation (2.5). For every left interval I such that $\mathbb{E}[\mu(I)X - Y] > 0$ by equation (4.3) we have $Z_n(I) \to +\infty$ a.s. and there are no more particles killed in I^c eventually as $n \to +\infty$. This implies the first statement of (3) and the statement (5). Conversely, if $\mathbb{E}[\mu(I)X - Y] \leq 0$ then by equation (2.5) we have $K_n(I^c) \sim -\frac{1}{2}\mathbb{E}[\mu(I)X - Y]$ almost surely as $n \to +\infty$. This implies (4). Finally, if $\mathbb{E}[\mu(I)X - Y] = 0$ then $Z_n(I) = 0$ i.o. almost surely, whence by $\mathbb{E}[X - Y] > 0$ it follows that $K_n(I^c) \to +\infty$ as $n \to +\infty$ almost surely. This implies the second statement of (3) (applied to the case $F(f_c)\mathbb{E}[X] - \mathbb{E}[Y] = 0$) and the statement (2) (applied to the case $F(f_c-)\mathbb{E}[X] - \mathbb{E}[Y] = 0$). $\qquad\square$

Proof of Theorem 2.2. (1) It is enough to consider $A \subseteq [0, 1] \cap [0, f_c)$ since there are no births in $A \setminus ([0, 1] \cap [0, f_c))$ almost surely. It follows immediately from Proposition 4.2(1); the uniform convergence comes from the inequality $Z_n(A) \leq Z_n(I)$ for all $A \subseteq I$ and $n \in \mathbb{N}$.

(2) By Proposition 4.2(1) we have that $Z_n/n \to +\infty$ almost surely. Recall that $f_{n,1}, f_{n,2}, \dots$ are i.i.d. We start by noting that

$$\sum_{i=1}^{\lfloor (n+1)/2 \rfloor} (\widetilde{X}_i - Y_i) \leq Z_n(A) \leq \sum_{i=1}^{\lfloor (n+1)/2 \rfloor} \widetilde{X}_i,$$

whence

$$\frac{\sum_{i=1}^{\lfloor (n+1)/2 \rfloor} (\widetilde{X}_i - Y_i)}{\sum_{i=1}^{\lfloor (n+1)/2 \rfloor} X_i} \leq \frac{Z_n(A)}{Z_n} \leq \frac{\sum_{i=1}^{\lfloor (n+1)/2 \rfloor} \widetilde{X}_i}{\sum_{i=1}^{\lfloor (n+1)/2 \rfloor} (X_i - Y_i)}.$$

Moreover, since a.s.

$$\frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} X_i} \overset{k \to \infty}{\to} 0,$$

it suffices to prove that a.s.

$$\frac{\sum_{i=1}^{k} \widetilde{X}_i}{\sum_{i=1}^{k} X_i} \overset{k \to \infty}{\to} \mu(A).$$

We recall that $\widetilde{X}_i$ is a sum of X_i Bernoulli random variables of parameter $\mu(A)$ and that the family of these Bernoulli variables is independent of the family $\{X_n\}_{n \in \mathbb{N}}$. Thus

$$\frac{\sum_{i=1}^{k} \widetilde{X}_i}{\sum_{i=1}^{k} X_i} = \frac{\mathcal{B}(\sum_{i=i}^{k} X_i, \mu(A))}{\sum_{i=1}^{k} X_i},$$

and we are therefore left to prove that a.s.

$$\frac{\sum_{i=1}^{N_n} W_i}{N_n} \overset{n \to \infty}{\to} \mu(A). \tag{4.7}$$

This follows easily from the fact that $\sum_{i=1}^{n} W_i / n \to \mu(A)$ almost surely. Denote by Ω_1 the set where there is convergence and, in addition, $N_n \to +\infty$ holds; clearly $\mathbb{P}(\Omega_1) = 1$. For all $\omega \in \Omega_1$, $\sum_{i=1}^{N_n(\omega)} W_i(\omega) / N_n(\omega)$ is a subsequence of the convergent sequence $\sum_{i=1}^{n} W_i(\omega)/n$.

(3) The a.s. convergence $Z_n/n \to \mathbb{E}[X - Y]/2$ as $n \to +\infty$ comes from Proposition 4.2. As for the second part, if A is an interval then the claim follows trivially by applying Proposition 4.2 to $Z_n(I)$ and Z_n. Let

$$\mathcal{B} := \{\cup_{i=1}^{k} [a_i, b_i] : a_1 < b_1 < a_2 < b_2 < \cdots < b_k, a_i, b_i \in \mathbb{Q}, \quad k = 1, 2, \ldots\}.$$

Since there are countable many intervals with rational endpoints,

$$\mathbb{P}\left(\frac{Z_n(B)}{Z_n} \to \mathbb{P}_\infty(B), \quad \forall B \in \mathcal{B}\right) = 1. \tag{4.8}$$

By the regularity of probability measures, it is easy to see that for every Borel set A, for every $\varepsilon > 0$ there exists sets $B_1, B_2 \in \mathcal{B}$, both depending on ε such that $B_1 \subset A \subset B_2$ and $\mathbb{P}_\infty(B_2 \setminus B_1) \leq \varepsilon$. Thus, if $Z_n(B)/Z_n \to \mathbb{P}_\infty(B)$ for every $B \in \mathcal{B}$, then for every $\varepsilon > 0$

$$\limsup_n \frac{Z_n(A)}{Z_n} \leq \limsup_n \frac{Z_n(B_2)}{Z_n} = \mathbb{P}_\infty(B_2) \leq \mathbb{P}_\infty(A) + \varepsilon,$$

$$\liminf_n \frac{Z_n(A)}{Z_n} \geq \liminf_n \frac{Z_n(B_1)}{Z_n} = \mathbb{P}_\infty(B_1) \geq \mathbb{P}_\infty(A) - \varepsilon,$$

so that equation (4.8) implies equation (2.4).

By the above arguments, it suffices to show the following: if $\mathbb{P}_n$, $\mathbb{P}_\infty$ are probability measures on $\mathbb{R}$, so that for every Borel set A, $\mathbb{P}_n(A) \to \mathbb{P}_\infty(A)$, then $\sup_t |F_n(t) - F_\infty(t)| =: \|F_n - F_\infty\|_\infty \to 0$, where F_n and F_∞ are the corresponding distribution functions. Let $\{x_i\}$ be the set of atoms of $\mathbb{P}_\infty$, $p_i := \mathbb{P}_\infty(\{x_i\})$ and let $H_\infty(t) = \sum_i p_i I_{(-\infty,t]}(x_i)$ be the distribution function of the measure $\sum_i p_i \delta_{x_i}$ (with total mass not larger than 1). Let, for every i, $p_i^n := \mathbb{P}_n(\{x_i\})$, by assumption $p_i^n \to p_i$. Let H_n be the distribution function of the measure $\sum_i p_i^n \delta_{x_i}$. Since $\mathbb{P}_n(\{x_1, x_2, \ldots\}) \to \mathbb{P}_\infty(\{x_1, x_2, \ldots\})$, from Scheffé's theorem

(see for instance [11, Lemma 5.10]), it follows that $\|H_n - H_\infty\|_\infty \to 0$. Since for every t, $F_n(t) \to F_\infty(t)$, we have that $(F_n(t) - H_n(t)) \to (F_\infty(t) - H_\infty(t))$ for every t. Moreover, since $\|H_n - H_\infty\|_\infty \to 0$ we have that $\lim_{t \to +\infty} H_n(t) \to \lim_{t \to +\infty} H_\infty(t)$ which, in turn, implies $\lim_{t \to +\infty}(F_n(t) - H_n(t)) \to \lim_{t \to +\infty}(F_\infty(t) - H_\infty(t))$ as $n \to +\infty$. The latter convergence along with the fact that function $F_\infty(t) - H_\infty(t)$ is continuous guarantees that the pointwise convergence implies uniform convergence. So, $\|F_n - F_\infty\|_\infty \leq \|H_n - H_\infty\|_\infty + \|(F_n - H_n) - (F_\infty - H_\infty)\|_\infty \to 0$. $\qquad\square$

References

[1] P. Bak, K. Sneppen, Punctuated equilibrium and criticality in a simple model of evolution. Phys. Rev. Lett. **74** (1993), 4083–4086.

[2] I. Ben-Ari, An empirical process interpretation of a model of species survival. Stochastic Process. Appl. **123** (2013), n. 2, 475–489. MR-3003360

[3] I. Ben-Ari, A. Matzavinos, A. Roitershtein, On a species survival model. Electron. Commun. Probab. **16** (2011), 226–233. MR-2788894

[4] W. Feller, An introduction to probability theory and its applications. Vol. I, Third edition, John Wiley & Sons Inc., New York-London-Sydney, 1968, xviii+509 pp. MR-0228020

[5] W. Feller, An introduction to probability theory and its applications. Vol. II, Second edition, John Wiley & Sons Inc., New York-London-Sydney, 1971, xxiv+669 pp. MR-0270403

[6] C. Grejo, F.P. Machado, A. Roldán-Correa, The fitness of the strongest individual in the subcritical GMS model. Electron. Commun. Probab. **21** (2016), n. 12, 5 pp. MR-3485381

[7] H. Guiol, F. P. Machado, R. B. Schinazi, A stochastic model of evolution. Markov Process. Related Fields **17** (2011), n. 2, 253–258. MR-2856242

[8] H. Guiol, F. P. Machado, R. B. Schinazi, On a link between a species survival time in an evolution model and the Bessel distributions. Braz. J. Probab. Stat. **27** (2013), n. 2, 201–209. MR-3028804

[9] R. Meester, D. Znameski, Limit behavior of the Bak-Sneppen evolution model. Ann. Probab. **31** (2003), 1986–2002. MR-2016609

[10] S. Michael, S. Volkov, On the generalization of the GMS evolutionary model. Markov Process. Related Fields **18** (2012), n. 2, 311–322. MR-2985139

[11] D. Williams, Probability with martingales. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991. xvi+251 pp. MR-1155402

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: http://www.imstat.org/publications/open.htm