

On a species survival model

Iddo Ben-Ari* Anastasios Matzavinos† Alexander Roitershtein‡

June 15, 2010; Revised March 18, 2011

Abstract

In this paper we provide some sharp asymptotic results for a stochastic model of species survival recently proposed by Guiol, Machado, and Schinazi.

1 Introduction and statement of results

Recently, Guiol, Machado, and Schinazi [7] proposed a new mathematical framework for modeling species survival which is closely related to the discrete Bak-Sneppen evolution model. In the original Bak-Sneppen model [3] a finite number of species are arranged in a circle, each species being characterized by its location and a parameter representing the *fitness* of the species and taking values between zero and one. The number of species and their location on the circle are fixed and remain unchanged throughout the evolution of the system. At discrete times $n = 0, 1, \dots$, the species with the lowest fitness and its two immediate neighbors update simultaneously their fitness values at random. The Bak-Sneppen evolution model is often referred to as an “ecosystem” because of the local interaction between different species. The distinguishing feature of the model, shown through numerical simulations in [3], is the emergence of self-organized criticality [1, 2, 6, 9] regardless the simplicity of the underlying evolution mechanism. The Bak-Sneppen model has attracted significant attention over the past few decades, but it has also been proven to be difficult for analytical study. See for instance [6] for a relatively recent survey of the model.

The asymptotic behavior of the Bak-Sneppen model, as the number of species gets arbitrarily large, was conjectured on the basis of computer simulations in [3]. It appears that the distribution of the fitness is asymptotically uniform on an interval $(f_c, 1)$ for some critical parameter f_c , the value of which is close to $2/3$ [1, 9].

Guiol, Machado, and Schinazi [7] were able to prove a similar result for a related model with a stochastically growing number of species. Their analysis is based on a reduction to the study of a certain random walk, which allows them to build a proof using well-known results from the theory of random walks. The main result of [7] is

*Dept. of Mathematics, University of Connecticut, Storrs, CT 06269, USA; iddo.ben-ari@uconn.edu

†Dept. of Mathematics, Iowa State University, Ames, IA 50011, USA; tasos@iastate.edu

‡Dept. of Mathematics, Iowa State University, Ames, IA 50011, USA; roiterst@iastate.edu

thus based on general properties of Markov chains, and suitable variations of the result can in principle be carried out to other similar models.

In this paper we focus on the model introduced in [7] as is (see also the recent work of the same authors [8]). Our aim is to elucidate the underlying mechanism responsible for the phenomenon described in [7] by sharpening the estimates that lead to the major qualitative statement therein. We proceed with a description of the Guiol, Machado, and Schinazi (GMS) model. In contrast to the Bak-Sneppen model, the number of species in the GMS model is random and changes in time, and only the species with the lowest fitness is randomly replaced. The local interaction between species is not considered in the GMS model, and therefore the spatial structure of the population is of no importance.

Let $p > \frac{1}{2}$ be given and denote $q = 1 - p$. Let \mathbb{Z}_+ denote the set of non-negative integers and let $X = (X_n : n \in \mathbb{Z}_+)$ be a discrete-time birth and death process with the following transition probabilities: from each state, X_n increases by 1 with probability p ; from each state different than 0, X_n decreases by 1 with probability $q = (1 - p)$; finally, at 0, X_n stays put with probability q . Thus X is a nearest-neighbor transient random walk on the integer lattice \mathbb{Z}_+ with holding times and reflection barrier at zero. A jump to the right represents birth of a new species, whereas a jump to the left represents death of an existing species. Thus X_n represents the number of *species alive* at time n . Throughout the paper we assume that $X_0 = 0$ with probability one.

When a new species is born, it is assigned a *fitness*. The fitness is a uniform $[0, 1]$ random variable independent on the fitness of all previously born species as well as of the path of the process X . When X jumps to the left, the species with the lowest fitness is eliminated. We remark that, in a different context, a similar model was considered by Liggett and Schinazi in [10].

Fix $f \in (0, 1)$. We examine the model by considering two coupled random processes, $L = (L_n : n \in \mathbb{Z}_+)$ (for *lower or left*) and $R = (R_n : n \in \mathbb{Z}_+)$ (respectively, for *right*), where L_n denotes the number of species alive at time n whose fitness is less than f while R_n denotes the number of the remaining species alive at time n .

Observe that L_n increases by 1 if X_n does and the newborn species has fitness less than f , and L_n decreases by 1 whenever X_n decreases by one and L_n is not zero. The value of L_n remains unchanged when either X_n increases by 1 and the newborn species has fitness at least f or X_n decreases by 1 and $L_n = 0$.

Notice that when it is not at zero, the process L evolves as a nearest-neighbor random walk with probability pf of jumping to the right, probability q of going to the left, and probability $1 - pf - q$ of staying put. When at zero, L_n jumps to the right with probability pf , and stays put with the complementary probability $1 - pf$. Thus L is itself a death and birth process. Since $p > q$, Markov chain L is positive recurrent if $pf < q$, null-recurrent if $pf = q$, and is otherwise transient. In what follows we will denote the critical value q/p of the parameter f by f_c .

It is shown in [7] that with probability one we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \cdot \#\{\text{species alive at time } n \text{ with fitness within } (a, b)\} = p(b - a), \quad (1)$$

for any interval $(a, b) \subset (f_c, 1)$. That is, the distribution of alive species with fitness higher than f_c approaches the uniform law on $(f_c, 1)$, while each species with fitness

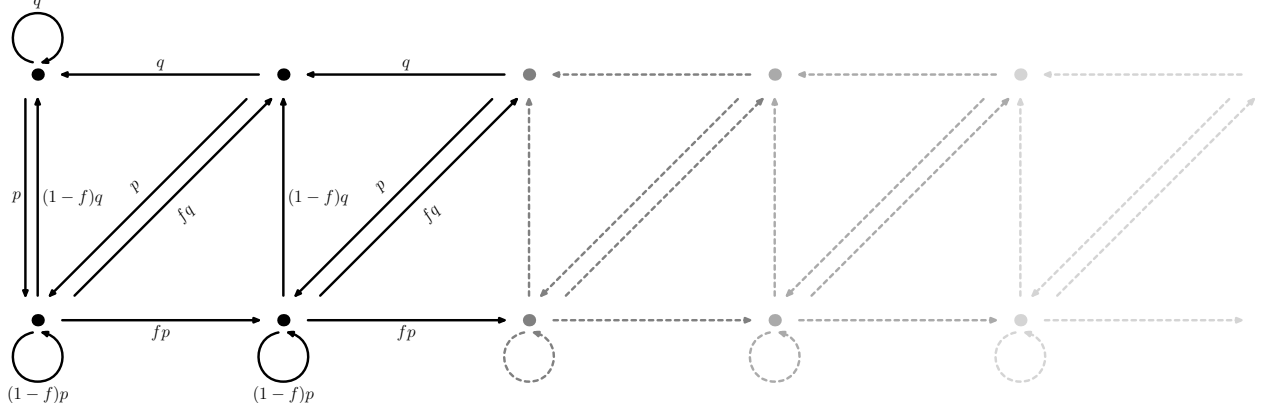


Figure 1: Transition probabilities for G

less than the critical value f_c disappears after a finite (random) time.

We sharpen (1) by proving the following theorem. Recall that $f_c = q/p$. Consider the process $B = (B_n : n \in \mathbb{Z}_+)$, where B_n is the total number of species born by time n with fitness at least f . Observe that B is a non-decreasing Markov chain (formed by sums of i.i.d. Bernoulli variables), which jumps one step up with probability $p(1-f)$ (equal to $p-q$ for $f = f_c$) and stays put with the complementary probability (which is equal to $2q$ for $f = f_c$). Hence by the law of large numbers, $\lim B_n/n = p(1-f)$ a.s., and (1) is immediate from the next result.

Theorem 1. *Suppose that $f = f_c$. Then*

1. $\limsup_{n \rightarrow \infty} \frac{B_n - R_n}{\sqrt{4qn \ln \ln n}} = 1, \quad \text{a.s.}$
2. $\frac{B_n - R_n}{\sqrt{2qn}} \Rightarrow |N(0, 1)|$, where $N(0, 1)$ denotes a mean-zero Gaussian random variable with variance one, and \Rightarrow stands for convergence in distribution.

2 Proof of Theorem 1

There is no loss of generality assuming that X is obtained recursively from an i.i.d. sequence of Bernoulli random variables $J = (J_n : n \in \mathbb{Z}_+)$ with $P(J_n = 1) = q$, $P(J_n = 0) = p$, as follows: $X_0 = 0$ and for $n \in \mathbb{Z}_+$ we have

$$X_{n+1} = X_n + (1 - J_n) - J_n(1 - s_n), \quad \text{where } s_n := \mathbf{1}_{\{X_n=0\}}. \quad (2)$$

Here and henceforth we use the standard notation $\mathbf{1}_A$ for the indicator of event A .

Let $G = (G_n : n \in \mathbb{Z}_+)$ be a Markov chain on $\mathbb{Z}_+ \times \{0, 1\}$ formed by the pairs $G_n = (L_n, J_n)$. Figure 1 illustrates the transition mechanism of G .

2.1 Reduction from $B_n - R_n$ to an occupation time of G

Let $\Delta = (\Delta_n : n \in \mathbb{Z}_+)$ be the process defined through $\Delta_n = B_n - R_n$. Notice that Δ_n increases by 1 if and only if R_n decreases by 1, and otherwise stays put. Since

$$\{R_{n+1} - R_n = -1\} = \{X_{n+1} - X_n = -1\} \cap \{L_n = 0\},$$

we have

$$\Delta_{n+1} - \Delta_n = \mathbf{1}_{\{X_{n+1} - X_n = -1\}} \mathbf{1}_{\{L_n = 0\}}.$$

Thus by (2), $\Delta_{n+1} - \Delta_n = (1 - s_n)J_n \mathbf{1}_{\{L_n = 0\}}$. Hence we have for $n \in \mathbb{N}$,

$$\Delta_n = \sum_{i=0}^{n-1} (1 - s_i)J_i \mathbf{1}_{\{L_i = 0\}} = \eta_{n-1} - \sum_{i=0}^{n-1} J_i s_i, \text{ where } \eta_n := \sum_{i=0}^n J_i \mathbf{1}_{\{L_i = 0\}}.$$

Observe that η_n is the occupation time (number of visits) of G at state $(0, 1)$ up to time n . Furthermore, since X is transient,

$$\sum_{i=0}^{\infty} s_i J_i \leq \sum_{i=0}^{\infty} s_i < \infty, \quad \text{a.s.}$$

Therefore, since L and consequently G are recurrent (and thus Δ_n is a non-decreasing sequence converging to $+\infty$ with probability one), it suffices to show that Theorem 1 holds with $\Delta_n = B_n - R_n$ replaced by η_n in its statement.

Excursion decomposition for the path of L We decompose the path of L into a sequence of successive excursions away from 0, each one begins at 0 and lasts until (but not including) the next time when L returns to 0 from 1. Set $V_0 = -1$ and, for $k \in \mathbb{N}$, let V_k be the total duration of the first k excursions of L from 0. That is

$$V_k = \inf\{n > V_{k-1} : L_n = 1, L_{n+1} = 0\}.$$

For $k \in \mathbb{N}$, define

$$\mu_k = \eta_{V_k} - \eta_{V_{k-1}} = \sum_{i=V_{k-1}+1}^{V_k} J_i \mathbf{1}_{\{L_i = 0\}}.$$

Notice that the excursions are independent and identically distributed. Therefore, $(\mu_k : k \in \mathbb{N})$ is an i.i.d. sequence. Let $N_m = \max\{k \in \mathbb{Z}_+ : V_k \leq m\}$, $m \in \mathbb{N}$. That is, N_m is the number of returns to 0 of the process L up to time m . Then for $m \in \mathbb{N}$ we have

$$\sum_{k=1}^{N_m} \mu_k \leq \eta_m < \sum_{k=1}^{N_m+1} \mu_k, \quad (3)$$

where we make the usual convention that $\sum_{k=1}^0 \mu_k = 0$.

We now compute $\mu := E(\mu_k)$ using the fact that the sequence of pairs $G_n = (L_n, J_n)$ forms a Markov chain, the transition mechanism of which is illustrated in Fig. 1. The value of μ is equal to the expected number of visits by this Markov chain to the state $(0, 1)$ during the period of time starting at the state $(0, 1)$ with probability q and at $(0, 0)$ with probability p , and lasting until G leaves the set $\{(0, 0), (0, 1)\}$. We thus have, using first step analysis,

$$\begin{aligned}\mu &= E(\mu_1) = P(J_0 = 1) \cdot (1 + \mu) + P(J_0 = 0, L_1 = 0) \cdot \mu + P(J_0 = 0, L_1 = 1) \cdot 0 \\ &= q(1 + \mu) + p(1 - f_c)\mu + pf_c \cdot 0 = q + \mu(1 - pf_c) = q + \mu p.\end{aligned}$$

Observe that once the Markov chain L is at zero, it will stay put until $J_k = 0$ and the fitness of the newborn particle is less than f_c . Hence μ_k is stochastically dominated by a geometric random variable with parameter $P(J_0 = 0, L_1 = 1) = pf_c = q$. In particular $\mu < \infty$, and hence the above identity implies $\mu = 1$. Consequently, using (3) and the law of large numbers, we obtain

$$\eta_n \sim N_n \text{ as } n \rightarrow \infty, \quad \text{a.s.} \quad (4)$$

Here and henceforth, $a_n \sim b_n$ as $n \rightarrow \infty$ for two sequences of real numbers $(a_n : n \in \mathbb{N})$ and $(b_n : n \in \mathbb{N})$ means, as usual, $\lim_{n \rightarrow \infty} a_n/b_n = 1$. In what follows, our plan is to prove a law of iterated logarithm and a central limit theorem for η_n by a reduction to the corresponding statements for the “inverse” V_n of N_n .

Reduction to a simple random walk With each excursion of L away from zero we can associate a *skeleton*, which is the path obtained from the excursion by omitting all transitions from a state of L to itself. The skeleton is an excursion of the simple (nearest-neighbor) symmetric random walk on \mathbb{Z}_+ with a reflection barrier at zero. Hence, if we let τ_k denote the length of the skeleton, then due to the choice of f_c it follows that τ_k has the same distribution as the time required for the simple (nearest-neighbor) symmetric random walk on \mathbb{Z} to get back to 0 starting from 0.

Recall that the holding time of L at zero during one excursion is a geometric random variable with the parameter $P(J_0 = 0, L_1 = 1) = pf_c = q$. Let h_k be the i.i.d. sequence of holding times at zero during successive excursions of L from zero. That is, $P(h_k = n) = p^{n-1}q$, $n \in \mathbb{N}$. In what follows we will use the notation $\text{Geom}(a)$ for the geometric distribution with parameter a (for instance, we could write $h_k \sim \text{Geom}(q)$). The time spent by L at each visit to a site is a geometric random variable, $\text{Geom}(2q)$ for sites different than 0 and $\text{Geom}(q)$ for 0.

Notice that the skeleton of the recurrent Markov process L is independent of the holding times at states visited during the excursion. Therefore, the length of a single excursion itself is a sum of one $\text{Geom}(q)$ random variable plus a sum of $\tau_k - 1$ independent $\text{Geom}(2q)$ random variables. If we replace h_k with a $\text{Geom}(2q)$, then the resulting modified “excursion time” becomes a sum of τ_k copies of a $\text{Geom}(2q)$ random variable. Let V'_m denote the total length of the first m excursions modified in this way. Clearly, $V_m \geq V'_m$. By the law of large numbers,

$$\lim_{m \rightarrow \infty} \frac{V_m - V'_m}{m} = E(h_1) - E(h'_1) = \frac{1}{2q}, \quad \text{a.s.}, \quad (5)$$

where h'_1 is a geometric random variable with parameter $2q$. Letting $T_m = \sum_{k=1}^m \tau_k$, we obtain

$$V'_m = \sum_{k=1}^{T_m} h''_k,$$

where $(h''_k : k \in \mathbb{N})$ is an i.i.d. sequence of random variables, each one distributed as $\text{Geom}(2q)$. Thus, by the law of large numbers,

$$V'_m \sim \frac{T_m}{2q} \text{ as } m \rightarrow \infty, \quad \text{a.s.} \quad (6)$$

Notice that T_m is distributed the same as the total length of the first m excursions from zero of a simple (nearest neighbor) symmetric random walk.

2.2 Completion of the proof: CLT and LIL for η_n

LIL for η_n We need the following result. Although the claim is a “folk fact”, we give a short proof for the sake of completeness.

Lemma 1.

$$\liminf_{m \rightarrow \infty} \frac{T_m}{m^2/(2 \ln \ln m)} = 1, \quad \text{a.s.}$$

Proof of Lemma 1. Let $S = (S_n : n \in \mathbb{Z}_+)$ denote the simple symmetric random walk on \mathbb{Z} . That is $S_0 = 0$ and

$$S_{n+1} = S_n + \zeta_n, \quad n \in \mathbb{Z}_+,$$

where $(\zeta_n : n \in \mathbb{Z}_+)$ is a sequence of i.i.d. random variables, taking values ± 1 with equal probabilities. Let $\gamma_0 = 0$ and define inductively $\gamma_{m+1} = \inf\{k > \gamma_m : S_k = m+1\}$ for $m \geq 0$. Let $\phi(x) = \sqrt{2x \ln \ln x}$ for $x > 0$. By the law of the iterated logarithm for S ,

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\phi(n)} = \limsup_{n \rightarrow \infty} \frac{S_{\gamma_n}}{\phi(\gamma_n)} = \limsup_{n \rightarrow \infty} \frac{n}{\phi(\gamma_n)}, \quad \text{a.s.}$$

Since $\phi^{-1}(k) \sim k^2/(2 \ln \ln k)$ as $k \rightarrow \infty$, we obtain

$$\liminf_{n \rightarrow \infty} \frac{\gamma_n}{n^2/(2 \ln \ln n)} = 1, \quad \text{a.s.} \quad (7)$$

Let $Y_m = \gamma_{m+1} - \gamma_m$, with the usual convention that the infimum over an empty set is $+\infty$. Observe that $\gamma_n = \sum_{i=0}^{n-1} Y_i$ and that $1 + Y_1$ is equal in distribution to the time length of an excursion of S away from zero. Thus the law of the sequence $(T_n : n \in \mathbb{N})$ is equal to the law of the sequence $(n + \gamma_n : n \in \mathbb{N})$, and we obtain

$$\liminf_{m \rightarrow \infty} \frac{T_m}{m^2/(2 \ln \ln m)} = \liminf_{n \rightarrow \infty} \frac{n + \gamma_n}{n^2/(2 \ln \ln n)} = 1, \quad \text{a.s.}$$

completing the proof of the lemma. □

Using the lemma along with (5) and (6), we obtain

$$\liminf_{m \rightarrow \infty} \frac{2qV_m}{m^2/(2 \ln \ln m)} = \liminf_{m \rightarrow \infty} \frac{2qV'_m}{m^2/(2 \ln \ln m)} = 1, \quad \text{a.s.}$$

Consequently, since N is the inverse sequence of V (that is, $V_{N_k} \leq k < V_{N_{k+1}}$), one can deduce from it (in the same way as (7) is derived from the usual LIL) that

$$\limsup_{k \rightarrow \infty} \frac{N_k}{\sqrt{4qk \ln \ln k}} = 1, \quad \text{a.s.}$$

Combining this with (4) completes the proof of the law of iterated logarithm for η_n .

CLT for η_n We now turn to the proof of the central limit theorem. The proof relies on a well-known limit theorem for a (properly normalized) random sequence T_m . More precisely, we have (see for instance [5, p. 394]):

$$\lim_{m \rightarrow \infty} E(e^{-\theta T_m/m^2}) = e^{-\sqrt{2\theta}}, \quad \theta \geq 0.$$

Therefore, it follows from (5) and (6) that

$$\lim_{m \rightarrow \infty} E(e^{-\theta V_m/m^2}) = \lim_{m \rightarrow \infty} E(e^{-\theta V'_m/m^2}) = e^{-\sqrt{\theta/q}}, \quad \theta \geq 0.$$

The function $\Phi_c(\theta) = e^{-c\sqrt{2\theta}}$, $\theta \geq 0$, with $c > 0$, is the Laplace transform of a positive stable law with index $1/2$ whose density function is given by (see for instance [5, p. 395])

$$\varphi_c(u) = \mathbf{1}_{\{u \geq 0\}} \frac{ce^{-c^2/2u}}{\sqrt{2\pi u^3}}.$$

We will use this formula with the parameter c equal to $c_* := \frac{1}{\sqrt{2q}}$. Observe that for all $k \in \mathbb{N}$ and $u > 0$,

$$P(N_k \leq u) = P_0(V_{\lfloor u \rfloor + 1} > k),$$

where $\lfloor u \rfloor$ stands for the integer part of u , that is $\lfloor u \rfloor = \max\{n \in \mathbb{Z}_+ : n \leq u\}$. Fix $s > 0$ and let $u = \sqrt{ks}$. Then, using the standard notation $o(1)$ to denote a sequence converging to zero when the underlying index k goes to infinity,

$$\begin{aligned} P(N_k \leq \sqrt{ks}) &= P(V_{\lfloor \sqrt{ks} \rfloor + 1} > k) \\ &= P\left(\frac{V_{\lfloor \sqrt{ks} \rfloor + 1}}{(\lfloor \sqrt{ks} \rfloor + 1)^2} > \frac{1}{s^2}(1 + o(1))\right) \xrightarrow[k \rightarrow \infty]{} \int_{s^{-2}}^{\infty} \varphi_{c_*}(u) du. \end{aligned}$$

Changing variables from u to $t = 1/\sqrt{u}$ in the last integral, we obtain

$$\lim_{k \rightarrow \infty} P(N_k \leq \sqrt{ks}) = \int_0^s \frac{2e^{-t^2/(2c_*^{-2})}}{\sqrt{2\pi c_*^{-2}}} dt.$$

Therefore, as $k \rightarrow \infty$, the random sequence N_k/\sqrt{k} converges weakly to the absolute value of a centered normal random variable with variance equal to $c_*^{-2} = 2q$. Combining this with (4) completes the proof of the central limit theorem for η_n . \square

Acknowledgement

We would like to thank the referee and associate editor for their comments and suggestions which allowed us to considerably improve the presentation of the paper.

References

- [1] P. Bak, *How Nature Works*, Copernicus, New York, 1996.
- [2] P. Bak, H. Flyvbjerg, and K. Sneppen, *Can we model Darwin?* *New Scientist* **1916** (1994), 36–39.
- [3] P. Bak and K. Sneppen, *Punctuated equilibrium and criticality in a simple model of evolution*, *Phys. Rev. Lett.* **74** (1993), 4083–4086.
- [4] P. Bak, C. Tang, and K. Wiesenfeld, *Self-organized criticality: an explanation of $1/f$ noise*, *Phys. Rev. Lett.* **59** (1987), 381–384.
- [5] R. Durrett, *Probability: Theory and Examples, 2nd edn.*, Duxbury Press, Belmont, CA, 1996.
- [6] A. Gillett, R. Meester, and P. van der Wal, *Maximal avalanches in the Bak-Sneppen model*, *J. Appl. Prob.* **43** (2006), 840–851.
- [7] H. Guiol, F. P. Machado, and R. B. Schinazi, *A stochastic model of evolution*, to appear in *Markov Processes and Related Fields*. Preprint is available at <http://arxiv.org/abs/0909.2108v2>.
- [8] H. Guiol, F. P. Machado, and R. B. Schinazi, *On a link between a species survival time in an evolution model and the Bessel distributions*. Preprint is available at <http://arxiv.org/abs/1102.2817v1>.
- [9] H. J. Jensen, *Self-Organized Criticality*, Cambridge University Press, 1998.
- [10] T. M. Liggett and R. B. Schinazi, *A stochastic model for phylogenetic trees*, *J. Appl. Probab.* **46** (2009), 601–607.
- [11] R. Meester and D. Znamenski, *Limit behavior of the Bak-Sneppen evolution model*, *Ann. Probab.* **31** (2003), 1986–2002.