Mathematical Biosciences 267 (2015) 53–60

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

A differential equation for the asymptotic fitness distribution in the Bak–Sneppen model with five species

Eckhard Schlemm

University College London Medical School, Gower Street, London WC1E 6BT, United Kingdom

ARTICLE INFO

Article history:
Received 9 March 2015
Revised 15 June 2015
Accepted 17 June 2015
Available online 3 July 2015

MSC:
37N25
92C42
60J05
62E20

Keywords:
Bak–Sneppen model
Evolutionary biology
Hypergeometric function
Markov chain
Steady-state distribution

ABSTRACT

The Bak–Sneppen model is an abstract representation of a biological system that evolves according to the Darwinian principles of random mutation and selection. The species in the system are characterized by a numerical fitness value between zero and one. We show that in the case of five species the steady-state fitness distribution can be obtained as a solution to a linear differential equation of order five with hypergeometric coefficients. Similar representations for the asymptotic fitness distribution in larger systems may help pave the way towards a resolution of the question of whether or not, in the limit of infinitely many species, the fitness is asymptotically uniformly distributed on the interval $[f_c, 1]$ with $f_c \gtrapprox 2/3$.

1. Introduction

The Bak–Sneppen (BS) model is an abstract representation of a biological system that evolves according to the Darwinian principles of random mutation and natural selection. It was introduced in [2] in the context of self-organized criticality in systems with spatial interactions.

Despite its simplicity, the BS model captures some of the features that are believed to be characteristic of evolving biological systems. In particular, it predicts evolutionary activity on all time scales with long periods of relative stasis interrupted by bursts of activities, referred to as *avalanches*. As a consequence of the absence of a characteristic time scale, evolutionary dynamics in the BS model display long-range dependence in both the temporal and the spatial domain. It is thus suitable as an abstract representation of systems in punctuated equilibrium, a concept that was introduced in [6] to explain the patterns observed in fossil records. For a more thorough discussion of these ideas, and for applications of the notion of punctuated equilibrium in other scientific disciplines we refer the reader to [12].

In addition to its usefulness in abstractly representing some key features of paleontology and macro-evolution, the BS model has also been employed to analyze the evolution of bacteria in a controlled, competitive environment. In a series of key experiments [14], Lenski and collaborators cultivated twelve initially identical populations of an *E. coli* strain over several years and conserved samples at regular time intervals. They then determined the relative fitness of the conserved samples by putting them into direct competition with a sample taken from the initial populations and measuring their relative growth rates. In [5] it was shown that the BS model with random mutations qualitatively reproduces some of the experimental results on relative bacterial fitness obtained in Lenski's long-term experimental evolution project [13]. Using an extended multi-trait variant, [3] extended the explanatory scope of the BS model to include experimental findings about the interplay of adaptation, randomness and history in bacterial evolution.

Informal definition of the BS model. The Bak–Sneppen model characterizes each species in a biological system by a numerical fitness value between zero and one, which represents its degree of adaptedness to its environment and changes as the species evolves. Further, each species is assumed to directly interact with exactly two other species, where it is left unspecified if such an interaction represents competition for resources, predator-prey relations or something different entirely. The Bak–Sneppen model can therefore be visualized as points on a circle, where each point stands for an ecological niche

---

E-mail address: eckhard.schlemm.13@ucl.ac.uk

http://dx.doi.org/10.1016/j.mbs.2015.06.011
0025-5564/

(or a species occupying that niche) and neighboring species interact with each other.

As time progresses, the fitness landscape evolves in accordance with the following rules, representing in an abstract way the principles of random mutation and natural selection: at each time step, the least adapted species, i.e. the one with the smallest fitness parameter, is removed from the system (becomes extinct) and its place is immediately taken by a new species whose fitness is initially modeled as an independent uniformly distributed random variable. In order to take into account the effect of this change on the local environment, the fitness parameters of the two species to either side of the least fit one are also reset to random values. This can be thought of as those two species themselves becoming extinct and superseded by new ones, or as them undergoing mutations in response to their neighbor becoming extinct.

Previous mathematical results. Despite its apparently easy definition, the BS model has withstood most attempts at mathematical analysis in the past. Partial results have been obtained, however, in the context of rank-driven processes [10,11] and mean-field approximations [4,7]. Early on it was conjectured based on simulations that the steady-state fitness distribution at a fixed site converges, in the limit of large populations, to a uniform distribution on the interval $[f_c, 1]$, where f_c is approximately equal to 0.667, but believed to be slightly larger than 2/3.

There is only a small number of mathematically rigorous result about the Bak–Sneppen model; in [15] it is shown that the steady-state fitness at a fixed site is bounded away from one in expectation, independent of the number of species in the system; A characterization of the limiting marginal fitness distribution, conditional on a set of critical thresholds, is given in [16] (see also [8,9]). In [18], the author proposes to compute the steady-state fitness distribution as the fixed point of the one-step transition equation and uses this method to describe the asymptotic fitness distribution for four species in terms of a compact rational function. In the same paper it is shown that one cannot find a similarly simple formula in the BS model with five species, and that the fitness distribution of a randomly selected species at steady-state in this case is not only not rational, but not even a hypergeometric function.

Our contribution. In this paper, we revisit the Bak–Sneppen model with five species. In Theorem 1, our main theorem, we establish a representation of the steady-state fitness distribution for five species in terms of the solution of an explicit differential equation with hypergeometric coefficients. This steady-state distribution encodes information about the fitness attributes of species in a system that has evolved for a long time. For instance, one can deduce from it how fit, on average, a randomly selected species from the population is expected to be; this is done in Corollary 1. Furthermore, since the steady-state distribution contains information about the joint fitness values of all species in the population, its knowledge allows to draw biologically relevant conclusions about qualitative properties of the system, such as the emergence of one or several dominant species, or the fragmentation of the eco-system into areas of different prevailing fitness. In our simple model, the symmetry of the initial configuration is preserved and no such phenomenon occurs. It is an interesting question whether in systems with a more complicated interaction between species, symmetry can be spontaneously broken.

We envisage that similar representations for the asymptotic fitness distribution in larger systems may help pave the way towards a resolution of the question of whether or not, in the limit of infinitely many species, the fitness is indeed asymptotically uniformly distributed on the interval $[f_c, 1]$ with $f_c \gtrapprox 2/3$. We speculate that the techniques developed in this paper can be generalized to analyze larger systems with more than five species. It seems plausible to expect that the asymptotic joint fitness distribution in such systems can still be characterized as the solution to a certain linear differential equation, even though the coefficients might no longer be hypergeometric functions.

2. Formalization and main result

We adopt the following formalization of the Bak–Sneppen model from [18]. Initially, all fitness parameters are independent uniformly distributed and after k evolutionary steps the state of the system is represented by the vector $\mathbf{f}_k \in [0, 1]^5$, where the ith component refers to the fitness of the ith species. The evolutionary dynamics of the system can be expressed formally by the equation

$$\mathbb{P}(\mathbf{f}_{k+1} \in A | \mathbf{f}_k = \mathbf{x}) = \int_A \mathbb{P}_{\mathbf{x}}(\mathrm{d}^5\xi), \quad \mathbf{x} \in [0,1]^5, \quad A \in \mathscr{B}([0,1]^5),$$

where the one-step transition kernel $\mathbb{P}_{\mathbf{x}}$ encodes the dynamics of the model and is given by

$$\mathbb{P}_{\mathbf{x}}(\mathrm{d}^5\xi) = \prod_{\mu \notin \{\nu-1,\nu,\nu+1\}} \delta_{x_\mu}(\mathrm{d}\xi_\mu)\mathrm{d}^3(\xi_{\nu-1}, \xi_\nu, \xi_{\nu+1}), \quad \nu = \operatorname{argmin} \xi.$$

Here, and in the following, all vector indices are taken modulo five. The sequence $\mathbf{f} = (\mathbf{f}_k)_k$ is a uniformly ergodic Markov chain with absolutely continuous marginal distributions with densities $g_k : [0,1]^5 \to \mathbb{R}^+$. This means that for any Borel set $A \in \mathscr{B}([0,1]^5)$,

$$\mathbb{P}(\mathbf{f}_k \in A) = \int_A g_k(\mathbf{x})\mathrm{d}^5\mathbf{x},$$

and that the random vectors $\mathbf{f}_k$ converge in distribution to a steady-state limit $\mathbf{f}_\infty$. Moreover, the k-step densities g_k satisfy the recursion

$$g_{k+1}(\mathbf{x}) = \sum_{\nu=1}^{5} \int_{[0,1]^3} \mathbf{1}_{\{\xi_2 < \min(\xi_1, \xi_3, \mathbf{x}_{]\nu[})\}} g_k\left(\mathbf{x}_{]\nu[_\xi}\right)\mathrm{d}^3\xi, \tag{1}$$

where the vectors $\mathbf{x}_{]\nu[} \in [0,1]^2$ and $\mathbf{x}_{]\nu[_\xi} \in [0,1]^5$ are obtained from $\mathbf{x}$ by dropping the νth, and $(\nu \pm 1)$th components, or replacing these components by the components of $\boldsymbol{\xi}$, respectively. Uniform ergodicity of the Markov chain $\mathbf{f}$ implies that the densities g_k converge uniformly to the density $g = g_\infty$ of the unique invariant distribution of $\mathbf{f}$, which we recognize as the steady-state fitness distribution. We also introduce the notation

$$F_{n,m}(x) = {}_2\mathrm{F}_1\left\{\frac{1}{3}\left(n + \mathrm{i}\sqrt{2}\right), \frac{1}{3}\left(n - \mathrm{i}\sqrt{2}\right); \frac{m}{3}; x\right\}, \quad n, m \in \mathbb{Z}, \quad x \in \mathbb{R},$$

where ${}_2\mathrm{F}_1$ denotes the Gauss hypergeometric function [1, Section 15.1] and $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. The following is the main result of the paper.

Theorem 1. *The limiting density $g = \lim_{k\to\infty} g_k$ is given by*

$$g(\mathbf{x}) = \mathbf{1}_{[0,1]^5}(\mathbf{x}) \sum_{\nu=1}^{5} q(\min\{x_\nu, x_{\nu+1}\}, \max\{x_\nu, x_{\nu+1}\}), \tag{2}$$

where $q(x, y) = \mathcal{G}'(1-x)\mathcal{B}_1'(1-y) + B_{\circ,0}(x)$. Here,

$$\mathcal{G}(x) = \frac{3}{2}F_{2,1}\{1/2\}F_{1,2}\left\{x^3/2\right\} + \frac{9}{8}xF_{4,5}\{1/2\}F_{2,4}\left\{x^3/2\right\}, \tag{3}$$

and the function $\mathcal{B}_1$ is the unique solution of the differential equation $\sum_{j=0}^{5} c_j(y)\mathcal{B}_1^{(j)}(y) = 0$ with boundary conditions

$$\mathcal{B}_1(1) = 1/5, \quad \mathcal{B}_1'(1) = 0, \quad \mathcal{B}_1''(1) = -1/5, \quad \mathcal{B}_1^{(3)}(1) = 1,$$
$$\mathcal{B}_1^{(4)}(1) = -18/5. \tag{4}$$

The coefficients $c_j(y)$, $j = 0, 1, \ldots, 5$, are hypergeometric functions given by

$$c_0(y) = \frac{18y^4}{(y^3+2)^2}\left[y\left(y^3 - 22\right)\mathcal{G}'(y) + \left(5y^3 - 14\right)\mathcal{G}(y)\right], \tag{5a}$$

$$c_1(y) = -yc_0(y), \tag{5b}$$