Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 123 (2013) 475–489

stochastic processes and their applications

www.elsevier.com/locate/spa

An empirical process interpretation of a model of species survival

Iddo Ben-Ari*

Department of Mathematics, University of Connecticut, 196 Auditorium Rd, Storrs, CT 06269-3009, United States

Received 24 August 2011; received in revised form 23 July 2012; accepted 18 September 2012
Available online 4 October 2012

Abstract

We study a model of species survival recently proposed by Michael and Volkov. We interpret it as a variant of empirical processes, in which the sample size is random and when decreasing, samples of smallest numerical values are removed. Micheal and Volkov proved that the empirical distributions converge to the sample distribution conditioned not to be below a certain threshold. We prove a functional central limit theorem for the fluctuations. There exists a threshold above which the limit process is Gaussian with variance bounded below by a positive constant, while at the threshold it is half-Gaussian.

Keywords: Species survival; Fitness; Central limit theorem; Empirical process

1. Introduction and statement of results

We study a generalization of the Guiol–Machado–Schinazi (GMS) model [5,2,6] that was recently proposed and analyzed by Michael and Volkov [8].

The model could be viewed as describing an ecosystem whose population size is given by a simple Markov chain on $\mathbb{Z}_+$. Each member of the ecosystem has a random "fitness" assigned at birth. When the population size decreases, the "least fit" members are eliminated. The population size process is driven by an IID sequence of $\mathbb{Z}$-valued random variables, $(I_n : n \in \mathbb{N})$. Starting with population size equal to 0 at time $n = 0$, at each time $n \in \mathbb{N}$, the population increases by I_n if $I_n \geq 0$, or decreases by the minimum between the present population size and $|I_n|$ if $I_n < 0$.

* Tel.: +1 860 2639773; fax: +1 860 486 4238.
E-mail address: iddo.ben-ari@uconn.edu.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2012.09.009

In the original GMS model, $I_n \in \{-1, 1\}$, that is the population size is modeled by a birth and death chain.

When $I_n \equiv 1$ for all n, the ecosystem at time n consists of n IID samples from a $U[0, 1]$ distribution, hence the immediate connection to empirical processes. In this model there are two additional ingredients. The first is the randomness of the sample size. This is not new, e.g. [9] and also, in a closely related context [1]. The former paper studies empirical processes in which the sample size is random and obeys a law of large numbers with positive speed. In the latter, among other things, the author proves a scaling limit for empirical distributions corresponding to sample size given by an increasing sequence of stopping times which are infinite with positive probability, conditioned to be finite. The second ingredient of the present model, and which appears to be new, is the mechanism according to which samples are discarded when the sample size decreases. This mechanism is responsible for criticality: the empirical distribution converges to the sample distribution conditioned not to drop below a certain threshold, in contrast to the classical Glivenko–Cantelli Theorem, where the empirical distributions converge to the sample distribution. A result of the same spirit holds in [1], due the conditioning. Furthermore, as we will show below, fluctuations from this distribution scale to a process which is discontinuous at the critical threshold. The process is Gaussian except at the critical threshold, where it is half-normal (the absolute value of a centered normal). This deviates from the "classical" Brownian Bridge scaling for empirical processes (e.g. [4, Theorem 14.3]), which is also the scaling limit in [9,1].

We turn to a formal description of the model. Let $I, I_1, \ldots$ be an IID sequence of $\mathbb{Z}$-valued random variables. We define the population size process X by letting

$$X_0 := 0, \qquad X_{n+1} := X_n + \max(I_{n+1}, -X_n), \quad n \in \mathbb{Z}_+.$$

This inductive formula gives the waiting time of the $n+1$-th customer in a $G/G/1$ queue, with I_{n+1} interpreted as the difference between the service time of n-th customer and interarrival time between n-th and $n+1$-th customer. However the main object of interest in the present model is the additional and intrinsic fitness structure, which does not translate naturally into queuing theory.

For $f \in [0, 1]$ and $n \in \mathbb{Z}_+$, let $L_n(f)$ denote the number of members of the population at time n whose fitness does not exceed f, and write $L(f) := (L_n(f) : n \in \mathbb{Z}_+)$ for the corresponding process. Here is an explicit construction. For $n \in \mathbb{Z}_+$, let $S_{n,+} := \sum_{0<j\le n}(I_j)_+$ and similarly, $S_{n,-} := \sum_{0<j\le n}(I_j)_-$, where here and henceforth we convene that summation over an empty index set has sum 0, and for a real number x, we define $x_+ := \max(x, 0)$, $x_- := (-x)_+ = -\min(x, 0)$. Let $U, U_1, \ldots$ be an IID sequence sampled from a $U[0, 1]$ distribution. For $f \in [0, 1]$ and $n \in \mathbb{Z}_+$, let

$$C_{n+1}(f) := \sum_{S_{n,+}<j\le S_{n+1,+}} \mathbf{1}_{[0,f]}(U_j),$$

that is, $C_{n+1}(f)$ represents the number of members of population born at time $n+1$ and whose fitness does not exceed f. As in the construction of X, we let

$$L_0(f) := 0, \qquad L_{n+1}(f) := L_n(f) + \max(C_{n+1}(f) - (I_{n+1})_-, -L_n(f)). \tag{1}$$

Since $C_{n+1}(f)$ and $(I_{n+1})_-$ are independent of $L_0, \ldots, L_n$, it follows that for each fixed f, $L(f)$ is a $\mathbb{Z}_+$-valued Markov chain. Note that $L_n(1) = X_n$.

We now provide an alternative construction which will be frequently utilized in the proofs. For $f \in [0, 1]$, let $S(f) := (S_n(f) : n \in \mathbb{Z}_+)$ denote the process consisting of the partial sums