

Dynamics of a Simple Evolutionary Process

Dietrich Stauffer
M. E. J. Newman

SFI WORKING PAPER: 2001-08-038

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

Dynamics of a simple evolutionary process

Dietrich Stauffer

Institute for Theoretical Physics, Cologne University, D-50923 Köln, Germany

M. E. J. Newman

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A.

(Dated: 31 July 2001)

Abstract

We study the simple evolutionary process in which we repeatedly find the least fit agent in a population of agents and give it a new fitness which is chosen independently at random from a specified distribution. We show that many of the average properties of this process can be calculated exactly using analytic methods. In particular we find the distribution of fitnesses at arbitrary time, and the distribution of the lengths of runs of hits on the same agent, the latter being found to follow a power law with exponent -1 , similar to the distribution of times between evolutionary events in the Bak–Sneppen model and models based on the so-called record dynamics. We confirm our analytic results with extensive numerical simulations.

I. INTRODUCTION

Is there any progress in the way humans live together? Yee [1] has recently suggested a simple model for the development of law-by-precedent, which assumes that bad decisions of legal courts have a higher tendency to be overruled by later and/or higher court decisions, thus improving the quality of the law over time. Yee's model can be thought of in general terms as an evolutionary process in which the least fit individuals in a population are successively removed and replaced by others [2]. In this paper we study this process analytically, providing some exact results and confirming these results with numerical simulations.

In the model proposed by Yee, a large number of agents $i = 1 \dots N$, representing biological species, court decisions, firms, etc. [3], each possess a fitness or quality $e_i \in [0, 1)$. Initially the e_i are uniformly distributed over their range. At each time-step, we find the individual that has the lowest fitness in the population and give it a new fitness which is chosen uniformly at random in the interval from zero to one. This model can be regarded as a simplified version of the Bak–Sneppen model of coevolution [2, 4], in which the least fit agent in a population and its neighbours on a d -dimensional lattice or other network are repeatedly removed and replaced with new agents with randomly chosen fitnesses. In the Bak–Sneppen model the agents typically represent species of organisms, and the replacement of neighbours arises as a result of interactions between species: host-parasite or prey-predator interactions, for instance [5]. In the case of court decisions and some other systems there is no strong case for such interactions and Yee thus omitted them, replacing instead only the element with the lowest fitness. Amongst other things this means that Yee's model does not show the self-organized criticality that is the principal focus of study in the Bak–Sneppen model, but Yee's model still has non-trivial behaviour.

Yee's model is also reminiscent of the so-called “record dynamics” [6, 7], which is the process which describes the pattern formed by the highest value seen so far in a sequence of random numbers. In both models, the interesting behaviour comes largely from the non-equilibrium nature of the dynamics.

It is worth noting that one does not need to choose fitnesses from the uniform distribu-

tion for the results given in this paper to be valid. Since the dynamics of Yee’s model is determined solely by the ranking of the agents relative to one another and is unaffected by their absolute fitness, one is free to chose numbers from any (normalizable) distribution and the results will be identical. We use uniform random numbers solely for ease of analysis and simulation.

II. ANALYTIC RESULTS

The basic form that the evolution of Yee’s model takes is clear. Suppose that $x(t)$ is the value of the lowest of the original distribution of fitnesses e_i which has not yet been touched at time t . Then the distribution of fitnesses above this value must be uniform, since all fitnesses are chosen at random from the uniform distribution. At time-step t , finding the lowest fitness in the population, we replace it with a random number which with probability $1 - x$ falls above x and hence is not the new lowest fitness. When this happens, the value of x increases by an amount which is given by $(1 - x)/N$ on average, and hence the rate at which x increases is

$$\frac{dx}{dt} = \frac{(1 - x)^2}{N}. \tag{1}$$

Defining a reduced time $\tau = t/N$, this has the solution

$$x(\tau) = \frac{\tau}{\tau + 1}. \tag{2}$$

Thus, as we would expect, the lowest fitness in the population increases monotonically and tends to 1 as $t \rightarrow \infty$.

This however does not tell the whole story. As the value of x increases, the chances of a new randomly chosen fitness falling above x decreases, and so at longer times it takes more attempts at replacing the fitness value e_i of an agent to find one which falls above x . Thus the dynamics of the model consists of “runs” of attempts at finding a new higher value for the fitness of the least fit agent.

Let us define a “run” to be the number of consecutive attempts to improve the fitness of a given agent until a new fitness is chosen which makes that agent no longer the least

fit individual in the population. The probability that a new run starts at time t is equal to the probability that the random number chosen at time $t - 1$ fell above x , which is $1 - x = 1/(\tau + 1)$. Thus the average length of a run at time t is $1/(1 - x) = \tau + 1$. The total number of runs from time 0 until a final time $t_f = N\tau_f$ is

$$R = \sum_{t=0}^{t_f-1} \frac{1}{t/N + 1} = N \int_0^{\tau_f} \frac{d\tau}{\tau + 1} = N \log(\tau_f + 1), \quad (3)$$

where the integral becomes exact in the limit of large system size with $\tau_f \gg 1$. Since each agent is equally likely to be the least fit on each run, the average number of runs which affect each agent is $R/N = \log(\tau_f + 1)$.

We can also calculate the complete distribution of the number k of runs which affect any given agent. Since, again, each agent is equally likely to be the least fit on each run, this is simply a binomial distribution

$$\begin{aligned} p_k &= \binom{R}{k} \left[\frac{1}{N} \right]^k \left[1 - \frac{1}{N} \right]^{R-k} \simeq \binom{R}{k} \frac{1}{N^k (\tau_f + 1)} \\ &\simeq \frac{[\log(\tau_f + 1)]^k}{k! (\tau_f + 1)}, \end{aligned} \quad (4)$$

where we have made use of the value of R from Eq. (3), and the last two equalities become exact for large N . When k is large, which it normally will be, one can further use Stirling's approximation $k! \simeq \sqrt{2\pi/k} (k/e)^k$ to write this as

$$p_k \simeq \frac{\sqrt{k/2\pi}}{\tau_f + 1} \left[\frac{e \log(\tau_f + 1)}{k} \right]^k. \quad (5)$$

The distribution of the lengths of runs can also be calculated exactly. At time t , the lowest fitness so far untouched is x , and the probability distribution of the lengths of runs is

$$p_n = x^{n-1} (1 - x), \quad (6)$$

which is correctly normalized, as can easily be verified. Since the probability that a new run starts at time t is $1 - x$, the overall probability distribution of the lengths of runs for the entire lifetime of the model is

$$P_n = \sum_{t=0}^{\infty} x^{n-1} (1 - x)^2. \quad (7)$$

Changing variables to τ again, using Eq. (2), and taking the limit of large N , the sum becomes an integral, and we find

$$P_n \propto \int_0^\infty \frac{\tau^{n-1}}{(\tau+1)^{n+1}} d\tau = \frac{1}{n}. \quad (8)$$

The lengths of the individual intervals between the evolution of one agent and the next have a non-stationary, monotonically increasing average as time passes. But Eq. (8) indicates that if one calculates an average over all times of the distribution of intervals, it follows a power-law with exponent -1 . This is reminiscent of the behaviour of the record dynamics, which also shows such an n^{-1} law when averaged over all times [8]. An n^{-1} law also appears in thermal barrier-crossing processes, such as those governing the times between evolutionary events in the Bak–Sneppen model. And similar power-law behaviour is seen in real-world macroevolution, where the distribution of the times between evolutionary events is clearly non-stationary [9]. If we equate evolutionary events with species extinction, then the logarithmic behaviour (3) in the cumulated number of events and the power-law distribution of the intervals between them (8) seen in the model are identical to those seen in the fossil record of extinctions [10].

Note that the distribution (8) is not integrable and therefore cannot be normalized. In any real case however—in simulations of the model for example—the distribution is truncated by finite-time effects and is perfectly normalizable.

The total number of times that the fitness of a given agent gets changed during the course of the system’s life also shows interesting behaviour. In theory, the number of runs in which each agent takes part, which is distributed according to the binomial distribution, Eq. (4), becomes increasingly sharply peaked about its mean as $\tau_f \rightarrow \infty$. Variation in the number of times a given agent is hit will then be a result of variation only in which particular set of runs affect each agent. Such sets are chosen independently at random from the distribution (8). Since this is a non-normalizable distribution, the sum representing the total number of hits violates the central limit theorem—its distribution will have a central portion which is approximately normally distributed, but there will be a tail of high values that, like the underlying distribution, goes as $1/n$, being the result of rare but statistically

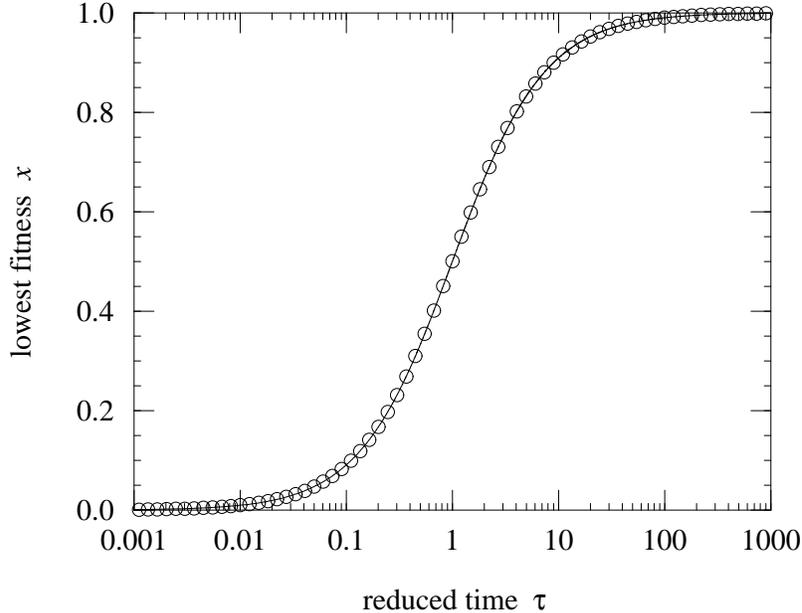


FIG. 1: Fitness x of the least fit member of the population as a function of the reduced time $\tau = t/N$. The points are simulation results for 10^9 time-steps with $N = 10^6$, and the solid curve is the exact solution in the large system-size limit, Eq. (2).

significant outliers sampled from the tail of the original distribution.

In practice, however, since the number of runs in which an agent takes part increases only as the logarithm of τ_f , Eq. (3), its value is small for all practical lifetimes of the system and hence shows substantial statistical fluctuation about its mean. In typical situations therefore, the variation in total number of hits an agent receives is dominated by these fluctuations, giving rise to a distribution which has an exponential tail. Only for exponentially long runs $\log(\tau_f + 1) \gg 1$ will the power-law behaviour be seen. Long though they are, none of the simulations we have performed fall into this category.

III. NUMERICAL RESULTS

Yee's model is straightforward to simulate on a computer, but the naive method of simulation, in which one searches through all agents on each time-step to find the least fit, is slow—it takes time $O(N)$ to perform the search and hence the entire simulation takes

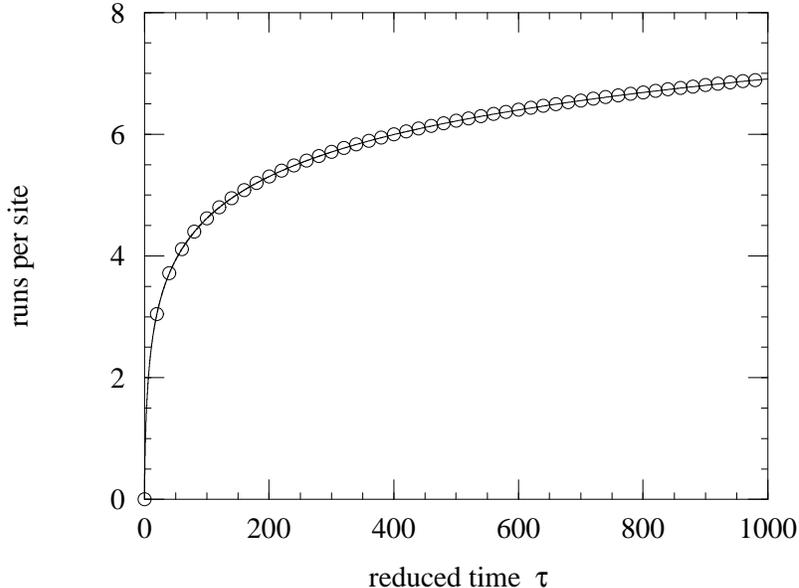


FIG. 2: Total number of “runs” per site as a function of reduced time. Points are simulation data for 10^9 time-steps with $N = 10^6$, and the solid curve is the analytic result, Eq. (3).

time $O(t_f N) = O(\tau_f N^2)$ to run for reduced time τ_f . A better approach is to store the agents’ fitnesses in a binary heap (a partially ordered binary tree with its smallest value at its root). This data structure allows us to find the agent with the lowest fitness in time $O(1)$, and add or remove agents in time $O(\log N)$, improving the running time of the algorithm to $O(\tau_f N \log N)$, which is fast enough for the simulation of quite large systems. For a description of the working and implementation of a binary heap, see for example Refs. 11 and 12.

In Fig. 1 we show simulation results for $x(\tau)$ as a function of τ , along with the exact solution (2), and as the figure shows, agreement between simulation and exact results is excellent. In Fig. 2 we show the total number of runs per site R/N as a function of τ during a simulation, along with the expected mean result, Eq. (3), and again the agreement is good. In Fig. 3 we show on logarithmic scales a histogram of the distribution of the lengths n of the intervals between evolutionary events. A straight-line fit to the data for smaller values of n indicates that the distribution is following a power-law with exponent -1.02 ± 0.01 , in good agreement with the predicted exponent of -1 , Eq. (8). For larger values of n , deviation

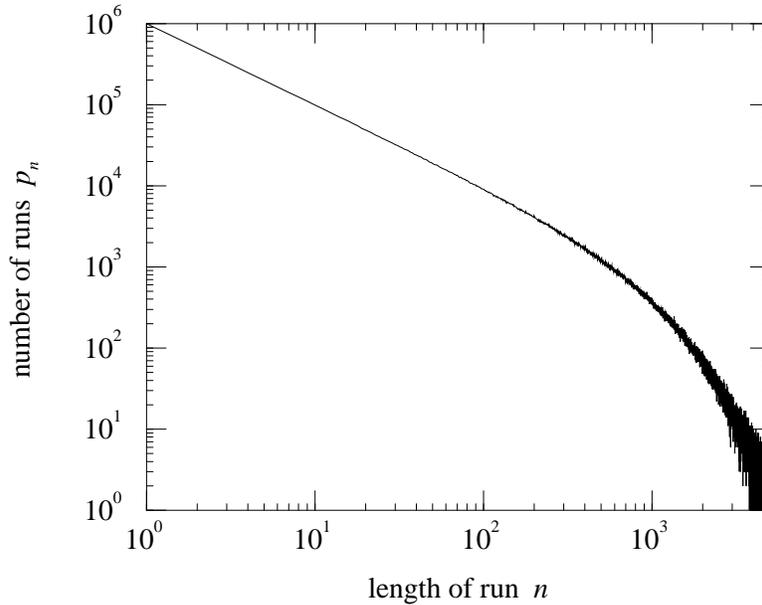


FIG. 3: Histogram of the lengths of runs in a simulation of the model for 10^9 time-steps with $N = 10^6$.

from the straight-line form is clearly visible in the region where n is of the order of τ_f or greater, where $\tau_f = 1000$ in this case.

IV. CONCLUSIONS

We have analysed the simple evolutionary process proposed by Yee [1] in which the agent with the lowest fitness in a large population is repeatedly removed and replaced with another having fitness chosen uniformly at random in a given interval. We have shown that many properties of the dynamics of this process can be calculated analytically. Using a numerical method employing a binary heap data structure, simulations of large realizations of the model are possible in reasonable time, and we have presented results from such simulations which confirm the analytic results.

ACKNOWLEDGEMENTS

The authors would like to thank Amnon Aharony for useful conversations. This work was supported in part by the National Science Foundation.

REFERENCES

- [1] K. K. Yee, “Introduction to spin and lattice models in the social sciences,” [nlin.A0/0106028](#).
- [2] P. Bak and K. Sneppen, “Punctuated equilibrium and criticality in a simple model of evolution,” *Phys. Rev. Lett.* **71**, 4083–4086 (1993); S. Boettcher and A. G. Percus, “Optimization with extremal dynamics,” *Phys. Rev. Lett.* **86**, 5211–5214 (2001); L. Furuberg, J. Feder, A. Aharony and J. Jøssang, “Dynamics of invasion percolation,” *Phys. Rev. Lett.* **61**, 2117–2120 (1988).
- [3] G. Cuniberti, A. Valleriani and J. L. Vega, “Effects of regulation on a self-organized market,” *Quantitative Finance* **1**, 332–335 (2001).
- [4] T. S. Ray and N. Jan, “Anomalous approach to the self-organized critical state in a model for ‘life at the edge of chaos’,” *Phys. Rev. Lett.* **72**, 4045–4048 (1994).
- [5] D. Stauffer and N. Jan, “Directed Bak-Sneppen model for food chains,” *Int. J. Mod. Phys. C* **11**, 147–151 (2000).
- [6] W. Feller, *An introduction to probability theory and its application*, 3rd edition, Wiley, New York (1967).
- [7] P. Sibani and P. B. Littlewood, “Slow dynamics from noise adaptation,” *Phys. Rev. Lett.* **71**, 1482–1485 (1993).
- [8] M. E. J. Newman, S. M. Fraser, K. Sneppen and W. A. Tozier, “Comment on ‘Self-organized criticality in living systems’,” *Phys. Lett. A* **228**, 202–204 (1997).
- [9] P. Sibani, M. R. Schmidt and P. Alstrøm, “Fitness optimization and decay of extinction rate through biological evolution,” *Phys. Rev. Lett.* **75**, 2055–2058 (1995).
- [10] M. E. J. Newman and G. J. Eble, “Decline in extinction rates and scale invariance in the fossil record,” *Paleobiology* **25**, 434–439 (1999).

- [11] D. E. Knuth, *The Art of Computer Programming, Vol. 2*, Addison–Wesley, Reading, Mass. (1973).
- [12] R. K. Ahuja, T. L. Magnanti and J. H. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, Englewood Cliffs (1993).