Probab. Theory Relat. Fields 133, 123–140 (2005)
Digital Object Identifier (DOI) 10.1007/s00440-004-0419-y

Jean Bérard

Genetic algorithms in random environments: two examples

Received: 12 June 2001 / Revised version: 8 November 2004 /
Published online: 6 June 2005 –

Abstract. We study the asymptotic behavior of two mutation-selection genetic algorithms in random environments. First, the state space is a supercritical Galton-Watson tree conditioned upon non-extinction and the objective function is the distance from the root. In the second case, the state space is a regular tree and the objective function is a sample of a tree-indexed random walk. We prove that, after n steps, the algorithms find the maximum possible value of the objective function up to a finite random constant.

1. Introduction

1.1. Motivation

Genetic algorithms are popular stochastic optimization methods, formally introduced by J. Holland [8] and based on a biological analogy. One seeks to maximize the *objective function* $f : S \to \mathbb{R}$, defined on the *state space* S. The mutation-selection genetic algorithms we consider evolve a finite population of particles in S under the following iterated two-stages procedure:

1. **Mutation:** each particle moves at random to a neighboring element of S, a convenient graph structure on S being given.
2. **Selection:** a new population of particles is created by resampling. Particles with low relative values of f tend to be eliminated, particles with large relative values of f are more often replicated.

Mutation allows for the discovery of new, possibly better, feasible solutions to the maximization of f, and selection concentrates the exploration of S around good, already discovered solutions. Note that the algorithms we study do not include the mating operator, since it is difficult to define it in our context. Moreover, genetic algorithms that include mating are in general harder to analyze than mutation-selection ones.

Despite the successes of this approach in the applications and numerous experimental studies, few rigorous mathematical results are available, that would explain how genetic algorithms work. In particular, the interaction between mutation, selection and population size remains poorly understood, even for simple examples. Among the exceptions is the paper by R. Cerf [4], which gives general asymptotic

J. Bérard: Ex-Laboratoire de Probabilités, Combinatoire and Statistique; Université Claude Bernard, Lyon I, France. e-mail: `jean.berard@univ-lyon1.fr`

convergence results in the context of Markov chains with rare transitions. See the papers by Y. Rabinovich and A. Wigderson [13], G. Rudolph [14] and J. Bérard and A. Bienvenüe [2], [3] where simple examples are worked out, and C. Mazza and D. Piau [11] for results in the infinite-population case.

In this paper, we study the asymptotic behavior of two mutation-selection genetic algorithms defined on trees in random contexts. First, the state space is a random supercritical Galton-Watson tree conditioned upon non-extinction, see for instance the paper [10] by R. Lyons, R. Pemantle and Y. Peres. The objective function $f(x)$ at a vertex x of the tree is the distance between x and the root of the tree. In the second case, the state space is a regular tree, and the objective function is a sample of a tree-indexed simple random walk, see [12].

Both situations display randomness, since either the graph structure of the state space or the objective function itself is random. In our opinion, this makes interesting the study of these simple optimization algorithms.

In both cases, the objective function is unbounded. We are looking for the asymptotic growth speed of the objective function values that the algorithm discovers, rather than for convergence rates to a finite maximum value. As explained by D. Aldous in [1], the study of the transient, in the Markov chain sense, behavior of infinite-state algorithms is intended as a toy model for the pre-equilibrium behavior of randomized optimization algorithms on large finite sets. Our second case is precisely the situation that Aldous analyses for the Metropolis algorithm.

A second motivation comes from the study of diffusion phenomena in random environments, see for instance the book by B. Hughes [9]. The biased random walk on Galton-Watson trees, studied in [10], corresponds to the Metropolis algorithm on these trees with our objective function $f(x)$. Thus, it is natural to ask for the behavior of genetic algorithms in the same context.

Our results show that, at least under some restrictions on the parameters, part of which are believed to be purely technical, the algorithms perform very well for the two examples considered. See below for a detailed formulation. Yet another motivation is the study of the approximation of infinite-population genetic algorithms by finite-population ones, that are the only computationally realistic algorithms. This is an important theoretical question since infinite-population genetic algorithms are often used to model the behavior of large population ones, see the paper by P. Del Moral, L. Kallel and J. Rowe [5], and the book by M. Vose [15]. From the results below, our two specific examples display such an approximation property, in a strong sense.

1.2. Description of the model and notations

We use proportional selection with Boltzmann weights, see [4], and mutation steps given by simple random walks on the state space. Instead of being kept constant as is usually the case, see [8], the population size p_n at time n, goes to infinity as n goes to infinity, at a possibly slow rate.

We use the same notations for the two cases, the case under consideration being clear from the context.

Galton-Watson case

Let $\mathbb{T}$ denote a random Galton-Watson tree, with root r. For $x \in \mathbb{T}$, $|x|$ is the tree-distance between r and x, and the objective function f is $f(x) = |x|$. The reproduction law of the Galton-Watson tree is q, and Q is a random variable of law q. We assume that q has bounded support, $Q \leq q^*$ say, and we set

$$\theta = \mathrm{E}(Q/(Q+1)).$$

Random objective function case

Fix $m \geq 2$, and let $\mathbb{T}$ denote a rooted regular $m-$ary tree, with root r. For every edge e of $\mathbb{T}$, choose $v(e) \in \{+1, -1\}$. Then, let $f(r) = 0$, and, for every vertex $x \neq r$ of $\mathbb{T}$,

$$f(x) = \sum_{e \in \{r \to x\}} v(e),$$

where $\{r \to x\}$ denotes the set of edges on the unique injective path from r to x. We assume that $(v(e))_e$ is a collection of i.i.d. random variables of law

$$\rho\, \delta_1 + (1-\rho)\, \delta_{-1}.$$

In both cases, we fix $\beta > 0$ and a deterministic sequence $(p_n)_n$ of positive integers. The genetic algorithm is a time-inhomogeneous Markov chain

$$X_n = (X_n^{(i)})_{1 \leq i \leq p_n},$$

such that $X_n \in \mathbb{T}^{p_n}$, with initial state $X_0 = (r, \ldots, r) \in \mathbb{T}^{p_0}$, defined by the following transitions.

1. **Mutation step:** $X_n \longrightarrow Y_n$
   Each particle $X_n^{(i)}$ evolves independently of the others, and performs one step of a simple random walk on $\mathbb{T}$, symmetric and to the nearest neighbors.
   In the random objective function case, the probability of staying at the same vertex is $\mu > 0$. In the Galton-Watson case, we assume that $\mu = 0$. The new positions are $Y_n^{(i)}$, $1 \leq i \leq p_n$.
2. **Selection step:** $Y_n \longrightarrow X_{n+1}$
   Each $X_{n+1}^{(i)}$ for $1 \leq i \leq p_{n+1}$ is chosen randomly and independently of the others in the set $\left\{Y_n^{(i)},\ 1 \leq i \leq p_n\right\}$, according to the probability law

$$\frac{1}{S_n'} \sum_{i=1}^{p_n} \exp\left(\beta f(Y_n^{(i)})\right) \delta_{Y_n^{(i)}}, \quad \text{where} \quad S_n' = \sum_{i=1}^{p_n} \exp\left(\beta f(Y_n^{(i)})\right).$$

The particles $X_n^{(1)}, \ldots, X_n^{(p_n)}$ compose the population at time n. Note that, at a given time, several particles may be located at the same vertex. Let N_n denote the counting measure of X_n, and Z_n the counting measure of the objective function values:

$$N_n = \sum_{i=1}^{p_n} \delta_{X_n^{(i)}}, \quad Z_n = \sum_{i=1}^{p_n} \delta_{f(X_n^{(i)})}.$$

For $x_1, \ldots, x_p \in \mathbb{T}$, let

$$f^*(x_1, \ldots, x_p) = \sup\{f(x_i)\,;\ 1 \le i \le p\}.$$

1.3. Statement of the results

We use the following assumptions.

- **A1**: $\lim_{n\to+\infty} p_n (\log n)^{-2} = +\infty$
- **A2**: $\lim_{n\to+\infty} p_{n+1}/p_n = 1$

Galton-Watson case

- **A3**: $\theta > 1/2$
- **A4**: $\exp\beta > 1/\theta$

Random objective function case

- **A5**: $m\rho \ge 1$
- **A6**: $\exp\beta > [(1-\mu)\,\rho\,]^{-1}(1+1/m)$
- **A7**: $\exp\beta > 1/\mu$

Theorem A. *Assume that A1, A2, A3 and A4 hold. Then, almost surely on the non-extinction of the underlying Galton-Watson tree $\mathbb{T}$, and almost surely on the realization of the algorithm,*

$$f^*(X_n) = n - O(1).$$

Moreover, the empirical law of $f(X_n)$ concentrates around its maximum value $f^*(X_n)$, in the following sense. There exists $c > 0$ such that, for all $a \ge 1$, and for all $n \ge k_a$, where k_a is random and almost surely finite,

$$p_n^{-1} Z_n([0, f^*(X_n) - a]) \le c\, \exp(-\beta a).$$

Theorem B. *Assume that A1, A2, A5, A6 and A7 hold. Then, almost surely on the random objective function f and on the realization of the algorithm,*

$$f^*(X_n) = n - O(1).$$

Moreover, the empirical law of $f(X_n)$ concentrates around its maximum value $f^*(X_n)$, in the sense of Theorem A:

$$p_n^{-1} Z_n((-\infty, f^*(X_n) - a]) \le c\, \exp(-\beta a).$$

We now comment upon these results and the hypotheses. Note that trivially $f^*(X_n) \le n$. Thus, in both cases, up to a finite random constant, the cloud of particles X_n contains the largest possible value of f after n steps, and the empirical distribution of the particles is concentrated, in the sense of an exponentially decreasing tail, around this maximum value. The interesting case in Theorem A is when $\mathrm{P}[Q = 0] > 0$, so that the tree contains many dead-ends, while having a positive probability of being infinite. Theorem A proves that, despite the many dead-ends in the tree, there is no slowdown of the genetic algorithm. On the other hand, let S_n denote the distance from the root of the Metropolis algorithm with temperature β after n steps (in our specific context, the Metropolis algorithm is nothing but the outward biased simple random walk defined in [10]). Let also S_n^* denote the supremum of p_n independent copies of S_n. Comparison of S_n with a biased simple random walk on $\mathbb{N}$ easily yields that, almost surely on non-extinction, $\mathrm{P}\left[S_n \geq (1-\varepsilon)n\right]$ decreases exponentially with n, for small enough ε. Since p_n increases only slowly (subexponentially) with n according to A2, $\mathrm{P}\left[S_n^* \geq (1-\varepsilon)n\right]$ also decreases exponentially with n. Moreover, for large enough β (see [10]), the Metropolis algorithm has zero asymptotic speed, that is, $\lim_{n\to+\infty} n^{-1}S_n = 0$ *a.s.* and it can easily be shown that, for such β, almost surely on non-extinction, for every $\varepsilon > 0$, $\mathrm{P}\left[S_n \geq \varepsilon n\right]$, whence $\mathrm{P}\left[S_n^* \geq \varepsilon n\right]$, decreases exponentially with n. We refer to the paper by A. Dembo, N. Gantert, Y. Peres and O. Zeitouni [6] for general large deviations results concerning biased random walks on Galton-Watson trees, although they do not cover the case $\mathrm{P}\left[Q = 0\right] > 0$. Hence, the genetic algorithm can be considered as more efficient than the Metropolis algorithm in this context. In the random objective function case, the algorithm finds the exponentially small part of the state space where f takes values of order n. In both cases, our proof of the efficiency of the interaction between mutation and selection relies upon the averaging effect produced by the growth of p_n.

Remark 1. The behavior stated in Theorems A and B is easily seen to be shared by the infinite-population versions of our algorithms. Since assumption A1 allows for a slow growth rate of the population size with time, our results may be interpreted as a strong approximation of the infinite-population behavior by finite-population algorithms. See [5] and [15] for an account on these questions.

Assumption A3 implies that the average number $\mathrm{E}(Q)$ of children under q, is strictly greater than one, so that the non-extinction of $\mathbb{T}$ occurs with positive probability and the conditioning in the theorem is trivially defined. We note that A3 is in general strictly stronger than the condition $\mathrm{E}(Q) > 1$, so our result does not include every supercritical Galton-Watson tree. However, the conclusion of Theorem A may hold as soon as $\mathrm{E}(Q) > 1$.

When A3 is satisfied, A4 is satisfied as soon as $\beta > \log(2)$, so β has not to be very large for the theorem to be valid. Whether the behavior for small values of β is different from the one stated in Theorem A is an open question.

Assumption A5 says that, given a vertex x, the average number of children y of x such that $f(y) = f(x) + 1$ is strictly greater than 1, so this condition is analogous to A3. In fact, there is a critical value $\rho_{\mathrm{cr}} \in (0, 1/m)$, see [12], such that, if $\rho < \rho_{\mathrm{cr}}$, there is almost surely only a finite number of vertices x such that $f(x) \geq -c$, for any c. When $\rho > \rho_{\mathrm{cr}}$, there exist infinite rays R in $\mathbb{T}$ and $c > 0$ such that $f(R(i)) \geq ci$ for all i. From A5, ρ is far away from ρ_{cr}, and indeed we have the stronger property that there exist infinite rays R in $\mathbb{T}$ and $c > 0$ such that $f(R(i)) \geq i - c$ for all i. However, when $m \geq 3$, this includes non-trivial optimization situations since, as soon as $\rho < 1/2$, the typical behavior of f along a ray is to go to $-\infty$ at a positive speed. Note that, even when $\rho > 1/2$, the result of Theorem B is not trivial: even though large positive values of f of order $(2\rho - 1)n$ are easy to find, values of order n are still exponentially rare at the $n-th$ level of the tree.

Theorems A and B still hold for more general mutation processes than those described above, with similar proofs, under suitable restrictions on the range of the parameters. For instance, we could allow the holding probability μ to be positive in the Galton-Watson case, or allow the walk to have unequal transition probabilities for different neighbours. However, the assumption $\mu > 0$ is needed in the random objective function case, to prove that there is a.s. an infinite number of record epochs (see Proposition 8 below).

The two cases considered have different structures, but the methods that we use for proving the results are so close that we give the full proof only in the Galton-Watson case. We sketch the slight modifications that are needed in the random function case.

2. Preliminaries on the Galton-Watson case

From now on, we assume that A1 to A4 are satisfied.

2.1. Additional notations and definitions

Let $M_n = Z_n(f^*(X_n))$ be the number of particles maintained by X_n at the top level. Define N'_n, Z'_n and M'_n in the same way as N_n, Z_n and M_n, where Y_n replaces X_n.

We use the standard construction for random diffusion processes in random environments. Thus, the Galton-Watson tree is progressively sampled as the algorithm explores it, rather than being fixed once for all. A standard result is that, if a property of the algorithm holds almost surely with respect to this construction, then it holds for almost every realization of the environment, and for almost every realization of the algorithm conditional on this environment. In this context, $\mathbb{T}_n$ denotes the subtree uncovered by the algorithm up to time n.

Remark 2. In all the proofs, we work on unconditioned Galton-Watson trees, and condition upon non-extinction only at the end.

The σ-algebra $\mathcal{F}_n$ is generated by X_k, Y_k and $\mathbb{T}_{k+1}$, for $0 \le k \le n-1$, and X_n. The σ-algebra $\mathcal{G}_n$ is generated by $\mathcal{F}_n$, $\mathbb{T}_{n+1}$ and Y_n.

Let $x_1, x_2, \dots, x_{L_n}$, with $L_n \ge 1$, denote the distinct vertices of the tree maintained by X_n at the level $f^*(X_n)$. Note that $L_n \le M_n$. Each x_i has a random, possibly zero, number z_i of children and, if $z_i \ge 1$, we denote these children by $x_i(1), \dots, x_i(z_i)$. Let $T_{n,i} = N_n(x_i)$ denote the number of particles maintained by the algorithm at x_i. Thus, $M_n = \sum_{i=1}^{L_n} T_{n,i}$. Let

$$T_n^* = \sup\{T_{n,i}\ ;\ 1 \le i \le L_n\}.$$

Definition 1. *Say that $H_n(\ell)$ holds if $T_n^* \le M_n/\ell$.*

Remark 3. $H_n(1)$ is the certain event.

2.2. Record epochs

Definition 2. *Call n a record epoch if $f^*(X_n) > f^*(X_p)$ for all $1 \le p \le n-1$.*

Record epochs are a fundamental tool. Indeed, in the construction above, fresh random decisions, that is independent from the past, are made concerning the number of children of $x_1, \dots, x_{L_n}$, at each record epoch n.

Proposition 4. *Conditional upon non-extinction of the tree, there exists almost surely an infinite number of record epochs.*

In other words, the sequence $f^*(X_n)$ is unbounded from above almost surely conditional on the tree being infinite.

Proof of Proposition 4. Assume that the mutation steps of the algorithm are defined as follows. Set $\Lambda = (q^* + 1)!$. First, for every $d \le q^* + 1$, fix a partition $\mathcal{P}_d$ of the set $\{1, \dots, \Lambda\}$ into d equal parts. Then, let d_i be the number of neighbors of $X_n^{(i)}$, $d_i \le q^* + 1$. We draw a random integer $u_i(n)$ according to the uniform law on $\{1, \dots, \Lambda\}$, and choose the next position $Y_n^{(i)}$ among the neighbors of $X_n^{(i)}$ according to the element of $\mathcal{P}_{d_i}$ to which $u_i(n)$ belongs.

Consider an arbitrary vertex $x \in \mathbb{T}$, a positive integer r, and a sequence $\xi(0), \dots, \xi(r-1)$ in $\{1, \dots, \Lambda\}$. Introduce the event

$$\Delta_n(x) = \{x \in X_n \text{ and } f(x) = f^*(X_n)\}.$$

On $\Delta_n(x)$, let i denote the smallest index $j \le p_n$ such that $X_n^{(j)} = x$. Say that the event $\Gamma_n(x)$ occurs if there exists a sequence $(i_k)_{0 \le k \le r}$ such that $i_0 = i$ and, for all $0 \le k \le r - 1$,

$$u_{i_k}(n+k) = \xi(k) \text{ and } X_{n+k+1}^{(i_{k+1})} = Y_{n+k}^{(i_k)}.$$

Finally, define

$$\Delta(x) = \{\Delta_n(x) \text{ occurs infinitely often}\}$$

and

$$\Gamma(x) = \{\Gamma_n(x) \text{ occurs infinitely often}\}.$$

A particle $Y_n^{(j)}$ such that $f(Y_n^{(j)}) \ge f^*(Y_n) - b$ has a probability greater than

$$1 - \left(1 - \exp(-\beta b) p_n^{-1}\right)^{p_{n+1}}$$

of being kept by the selection step, conditional on $\mathcal{G}_n$. After r mutation steps, the distance between the current maximum and a particle initially located on the maximum cannot exceed $2r$ (in the worst case, the particle performs r downward steps, and the maximum performs r upward steps). Moreover, for a given mutation step, $\xi(k)$ is drawn with probability Λ^{-1}. Hence, on $\Delta_n(x)$,

$$\mathrm{P}[\Gamma_n(x) \mid \mathcal{F}_n] \ge \Lambda^{-r} \left(1 - \left(1 - \exp(-\beta 2r) p_n^{-1}\right)^{p_{n+1}}\right)^r.$$

This expression being bounded away from zero independently from n, by Lemma 13,

$$\mathrm{P}[\Delta(x) \setminus \Gamma(x)] = 0. \tag{1}$$

Call B the event that $f^*(X_n)$ is bounded from above. If B occurs, there is at least one vertex $x \in \mathbb{T}$ such that $\Delta(x)$ occurs. Hence, according to (1) and Lemma 13, for every finite sequence $\xi(0), \dots, \xi(r-1)$, $\Gamma(x)$ occurs almost surely on B. But the non-extinction of $\mathbb{T}$ implies that, for all $C > 0$, there is a positive integer r and a path of length r connecting x to another vertex at distance greater than C from the root. This path is associated with at least one sequence of mutation decisions $\xi(0), \dots, \xi(r-1)$. Since this holds for all C, we get that, conditional on non-extinction, B has zero probability. □

3. Proof in the Galton-Watson case

The theorem is a simple consequence of Propositions 6 and 7. The technical lemmas needed are stated in section 5. Lemma 5 below contains our basic estimates on the one-step transitions of the algorithm. We start with the following definition.

Definition 3. *For $\lambda > 0$, define the event $D_1(n)$ by:*

- $f^*(X_{n+1}) = f^*(X_n) + 1$
- $M_{n+1} \geq \lambda p_{n+1}$ *if* $M_n \geq \lambda p_n$ *and* $M_{n+1} \geq \nu^2 M_n$ *if* $M_n \leq \lambda p_n$

For $\ell \geq 1$, set

$$D_2(n, \ell) = D_1(n) \cap H_{n+1}(\nu\, \ell).$$

Throughout the proofs, we use a pair of parameters (ψ, ν) chosen as follows. Let $\psi \in (1/\theta, \exp(\beta))$ (from A4, $\theta \exp(\beta) > 1$). Let $\nu > 1$ such that $\nu^2 < \psi\theta$ and $\nu < 2\theta$ (from A3, $\theta > 1/2$).

Lemma 5. *There exist $\lambda > 0$, $K > 0$, $n_0 \geq 1$ and $g > 0$, such that, for all $n \geq n_0$,*

- *if n is a record epoch,*

$$\mathrm{P}\left[D_1(n) \mid \mathcal{F}_n \right] \geq g,$$

- *if n is a record epoch, if $M_n \geq K$, and if $H_n(\ell_n)$ occurs with $\ell_n \leq (M_n)^{1/2}$, then*

$$\mathrm{P}\left[D_2(n, \ell_n) \mid \mathcal{F}_n \right] \geq 1 - C_1 \exp\left(-C_2 (M_n)^{1/2}\right) - \exp\left(-C_3 \ell_n\right).$$

(recall that the definitions of $D_1(n)$ and $D_2(n, \ell)$ involve a parameter λ.)

Proof of Lemma 5. Set $B_n = \{n \text{ is a record epoch}\} \cap H_n(\ell_n)$. We put $\ell_n = 1$ for the proof of the first assertion of the lemma. Let $t_n = M_n/\ell_n$. Let $\epsilon \in (0, 1)$, whose value will be fixed later. Denote by $C_r(expression)$, any positive constant that depends only on $expression$ and on the law q.

Random sampling of the tree

Conditionally to $\mathcal{F}_n$ and B_n, n is a record epoch and the numbers of children z_i, $1 \leq i \leq L_n$ of the vertices x_i, $1 \leq i \leq L_n$ are i.i.d. random variables with law q. Let

$$J_n = \sum_{i=1}^{L_n} T_{n,i}\, z_i/(z_i + 1), \qquad E_1 = \{J_n \geq (1 - \epsilon)\theta M_n\}.$$

From Lemma 10, on B_n,

$$\mathrm{P}\left[E_1^c \mid \mathcal{F}_n\right] \le \exp\left(-C_1(\epsilon)M_n/T_n^*\right).$$

Hence,

$$\mathrm{P}\left[E_1^c \mid \mathcal{F}_n\right] \le \exp\left(-C_1(\epsilon)\ell_n\right) = e_1.$$

Mutation

Conditionally to $\mathcal{F}_n$ and $\mathbb{T}_{n+1}$, the number of particles $I_1, \dots, I_{L_n}$ that move by mutation from each vertex x_i to the children of x_i are independent random variables, with respective binomial laws $\mathcal{B}(T_{n,i}, z_i/(z_i+1))$. Let

$$E_2 = \{M_n' \ge (1-\epsilon)^2\theta M_n,\, f^*(Y_n) = f^*(X_n) + 1\}.$$

The fact that $I_1 + \cdots + I_{L_n} \ge (1-\epsilon)^2\theta M_n$ implies that $f^*(Y_n) = f^*(X_n)+1$ and that $M_n' = I_1 + \cdots + I_{L_n}$ since $(1-\epsilon)^2\theta M_n > 0$. Thus, E_1 and $I_1 + \ldots + I_{L_n} \ge (1-\epsilon)J_n$ imply E_2. By Lemma 11, on E_1,

$$\mathrm{P}\left[E_2^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \le \exp\left(-C_2(\epsilon)J_n\right)$$

so that on E_1,

$$\mathrm{P}\left[E_2^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \le \exp\left(-C_3(\epsilon)M_n\right) = e_2$$

Let us define the set of indices $\mathcal{A}_n$ by:

$$\mathcal{A}_n = \{1 \le i \le L_n \mid T_{n,i} \ge (1/2)t_n, z_i \ge 1\}.$$

Fix $i \in \mathcal{A}_n$ and $1 \le j \le z_i$. The number $N_n'(x_j)$ of particles that move by mutation from x_i to $x_i(j)$, follows, conditionally to $\mathcal{F}_n$ and $\mathbb{T}_{n+1}$, a binomial law $\mathcal{B}(T_{n,i}, (z_i+1)^{-1})$. Let

$$E_3(i,j) = \{N_n'(x_i(j)) \le (1+\epsilon)T_{n,i}(z_i+1)^{-1}\}.$$

We have by Lemma 12

$$\mathrm{P}\left[(E_3(i,j))^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \le \exp\left(-C_4(z_i,\epsilon)T_{n,i}\right).$$

Using the fact that $z_i \le q^*$, and that $T_{n,i} \ge (1/2)t_n$ since $i \in \mathcal{A}_n$, we get

$$\mathrm{P}\left[E_3(i,j)^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \le \exp\left(-C_5(\epsilon)t_n\right).$$

Let

$$E_3 = \bigcap_{i \in \mathcal{A}_n} \bigcap_{j=1}^{z_i} E_3(i,j).$$

The above intersection runs over at most $q^* M_n$ events. Hence

$$\mathrm{P}\left[E_3^c \mid \mathcal{F}_n, \mathbb{T}_{n+1}\right] \le q^* M_n \exp\left(-C_5(\epsilon)t_n\right) = e_3.$$

Selection

Define E_4 by:

- on $\{M_n > \lambda p_n\}$,

$$M_{n+1} \geq \psi\theta(1-\epsilon)^4 \lambda p_{n+1}, \quad f^*(X_{n+1}) = f^*(Y_n),$$

- on $\{M_n \leq \lambda p_n\}$,

$$M_{n+1} \geq \psi\theta(1-\epsilon)^4 M_n, \quad f^*(X_{n+1}) = f^*(Y_n).$$

From now on, we assume that ϵ is small enough so that: $\psi\theta(1-\epsilon)^4 \geq \nu^2$, and we choose λ as in Lemma 9. By Lemma 9, conditionally to $\mathcal{G}_n$, the number $Z_{n+1}(f^*(Y_n))$ of particles at level $f^*(Y_n)$ contained in X_{n+1} follows a binomial law $\mathcal{B}(p_{n+1}, \mu_{n+1})$, with

$$\mu_{n+1} \geq \psi \inf(\lambda, M'_n p_n^{-1}) = h.$$

Let

$$G = \{Z_{n+1}(f^*(Y_n)) \geq (1-\epsilon) h p_{n+1}\}.$$

Since $h > 0$, G implies that $M_{n+1} = Z_{n+1}(f^*(Y_n))$. Moreover, one can easily check that for n sufficiently large (that is, such that $p_{n+1}/p_n \geq 1-\epsilon$)), $G \cap E_2 \subset E_4$. On the other hand, we have $h \leq \psi\lambda$, and by Lemma 12, on E_2,

$$\mathrm{P}\left[E_4^c \;\middle|\; \mathcal{G}_n \right] \leq \exp\left(-C_6(\lambda, \epsilon) M_n\right) = e_4.$$

Conditionally to $\mathcal{G}_n$, M_{n+1} and $f^*(X_{n+1})$, the number $N_{n+1}(y)$ of particles put by selection on a vertex y at level $f^*(X_{n+1})$ follows a binomial law $\mathcal{B}(M_{n+1}, N'_n(y)$ $(Z'_n(f^*(X_{n+1})))^{-1})$. Assume that E_4 occurs, so that $f^*(X_{n+1}) = f^*(Y_n)$. Hence $Z'_n(f^*(X_{n+1})) = M'_n$. For $i \in \mathcal{A}_n$ and $1 \leq j \leq z_i$, E_3 implies that $N'_n(x_i(j)) \leq (1+\epsilon)(1/2)T_n^*$, and E_2 implies that $M'_n \geq (1-\epsilon)^2\theta M_n$. Thus, on $B_n \cap E_2 \cap E_3$:

$$N'_n(y)(M'_n)^{-1} \leq (2\theta)^{-1}(1+\epsilon)(1-\epsilon)^{-2}(\ell_n)^{-1}.$$

Let

$$E_5(i,j) = \{N_{n+1}(x_i(j))(M_{n+1})^{-1} \leq (2\theta)^{-1}(1+\epsilon)^2(1-\epsilon)^{-2}(\ell_n)^{-1}\}.$$

From now on, we assume that ϵ is small enough so that $(2\theta)^{-1}(1+\epsilon)^2(1-\epsilon)^{-2} \leq \nu^{-1}$. By lemma 12, on $B_n \cap E_2 \cap E_3 \cap E_4$,

$$\mathrm{P}\left[E_5(i,j)^c \;\middle|\; \mathcal{G}_n, M_{n+1}, f^*(X_{n+1}) \right] \leq \exp\left(-C_8(\epsilon) M_{n+1} \ell_n^{-1}\right).$$

Note that, on E_4, we have (easily) $M_{n+1} \geq \lambda M_n$ since $\lambda < 1$. Let

$$E_5 = \bigcap_{i \in \mathcal{A}_n} \bigcap_{j=1}^{z_i} E_5(i,j).$$

On $B_n \cap E_2 \cap E_3 \cap E_4$,

$$\mathrm{P}\left[E_5^c \;\middle|\; \mathcal{G}_n, M_{n+1}, f^*(X_{n+1}) \right] \leq q^* M_n \exp\left(-C_9(\epsilon, \lambda) M_n / \ell_n\right) = e_5.$$

When y's parent is one of the x_i, $i \notin \mathcal{A}_n$, we have, $N_n'(y) \le (1/2)t_n$. Let

$$E_6(i,j) = \{N_{n+1}(x_i(j))(M_{n+1})^{-1} \le (1/2)\theta^{-1}(1+\epsilon)(1-\epsilon)^{-2}(\ell_n)^{-1}\}$$

and

$$E_6 = \bigcap_{i\notin\mathcal{A}_n} \bigcap_{j=1}^{z_i} E_6(i,j).$$

Using the same argument as for E_5, we get that, on $E_2 \cap E_4$,

$$\mathrm{P}\left[E_6^c \mid \mathcal{G}_n, M_{n+1}, f^*(X_{n+1})\right] \le q^* M_n \exp\left(-C_{11}(\epsilon,\lambda)M_n/\ell_n\right) = e_6.$$

Set $U_n = E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5 \cap E_6$ and $W_n = E_1 \cap E_2 \cap E_4$.

Let us summarize the properties of X_{n+1} contained in $B_n \cap U_n$.

- $f^*(X_{n+1}) = f^*(X_n) + 1$ and so $n+1$ is a record epoch.
- $M_{n+1} \ge \lambda p_n$ if $M_n \ge \lambda p_n$, and $M_{n+1} \ge \nu^2 M_n$ if $M_n \le \lambda p_n$
- $T_{n+1}^* M_{n+1}^{-1} \le \nu^{-1}\ell_n^{-1}$ and so $H_{n+1}(\nu\ell_n)$ occurs

The two first properties are also consequences of $B_n \cap W_n$. Note that the e_i are measurable with respect to $\mathcal{F}_n$, and that $e_1, e_2, e_4 < 1$. Moreover, assuming $\ell_n \le M_n^{1/2}$ and $M_n \ge K$ for K large enough, we get that: $e_7 = e_1 + e_2 + e_3 + e_4 + e_5 + e_6 < 1$. Indeed, $e_1 \le \exp(-C_1(\epsilon)\ell_n)$, and all the other terms are bounded from above by an expression of the form $C_{12}M_n \exp(-C_{13}(\epsilon,\lambda)M_n/\ell_n)$. The inequality $\ell_n \le (M_n)^{1/2}$ entails that this expression is itself bounded from above by $C_{12}M_n\exp(-C_{13}(\epsilon,\lambda)M_n^{1/2})$, hence by $C_{14}\exp(-C_{15}(\epsilon,\lambda)M_n^{1/2})$, and this last expression goes to zero as $M_n \to +\infty$.

Conditioning $\mathrm{P}\left[U_n \mid \mathcal{G}_n, M_{n+1}, f^*(X_{n+1})\right]$, we get that, if $\ell_n \le (M_n)^{1/2}$, on $B_n \cap \{M_n \ge K\}$,

$$\mathrm{P}\left[U_n \mid \mathcal{F}_n\right] \ge (1-e_6-e_5)(1-e_4)(1-e_3-e_2)(1-e_1) \ge 1-e_7,$$

and, on B_n, without restrictions on ℓ_n or M_n,

$$\mathrm{P}\left[W_n \mid \mathcal{F}_n\right] \ge (1-e_4)(1-e_2)(1-e_1).$$

□

Proposition 6 below is the key result in our proof of the main theorem. The idea of its proof is the following: starting at a record epoch, M_n has a positive (conditional) probability of growing exponentially fast until $M_n \ge \lambda p_n$ and then to remain greater than λp_n forever. Such a regular growth is made possible by maintaining many different vertices at level $f^*(X_n)$, no vertex containing too many particles, so that an averaging phenomenon occurs.

Proposition 6. *Conditional on the non-extinction of the tree, there exists $\lambda > 0$ such that there exists almost surely a time k such that, for all $n \ge k$, $M_n \ge \lambda p_n$ and $f^*(X_n) = f^*(X_{n-1}) + 1$.*

Proof of Proposition 6. We take λ, $K \geq 1$, n_0 as they were defined in the proof of Lemma 5. Let $n_1 \geq n_0$ be such that, for all $n \geq n_1$, $\lambda p_n \geq K$, $p_{n+1}/p_n \leq \nu^2$. (Such an n_1 exists thanks to assumptions A1 and A2.) Assume throughout that r is a record epoch greater than n_1. Let r_1 be such that $\nu^{2r_1} \geq K$ (remember that $\nu > 1$). Let

$$R_1 = \{r + r_1 \text{ is a record epoch}, M_{r+r_1} \geq K\}.$$

Note that

$$\bigcap_{i=r}^{r+r_1-1} D_1(i) \subset R_1.$$

By Lemma 5,

$$\mathrm{P}\,[\,R_1 \mid \mathcal{F}_r] \geq g^{r_1}.$$

Let $r_2 = \inf\{k \geq 0 \mid \nu^{2k} K \geq \lambda p_{r+r_1+k}\}$ (r_2 is well-defined thanks to A2.) Let

$$R_2 = \{r + r_1 + r_2 \text{ is a record epoch}\} \cap \{M_{r+r_1+r_2} \geq \lambda p_{r+r_1+r_2}\} \cap H_{r+r_1+r_2}(\nu^{r_2}).$$

We set $D_3(r, r_1, -1) = \Omega$ and, for $k \geq 0$,

$$D_3(r, r_1, k) = \bigcap_{i=r+r_1}^{r+r_1+k} D_2(i, \nu^{i-r-r_1}).$$

According to the definition of r_2:

$$D_3(r, r_1, r_2 - 1) \cap R_1 \subset R_2.$$

For $0 \leq k \leq r_2 - 1$,

$$\mathrm{P}\left[\, D_3(r, r_1, k) \,\middle|\, \mathcal{F}_{r+r_1+k}\right] = \mathrm{P}\left[\, D_2(r + r_1 + k, \nu^k) \,\middle|\, \mathcal{F}_{r+r_1+k}\right] \mathbf{1}_{D_3(r,r_1,k-1)}.$$

On $D_3(r, r_1, k-1) \cap R_1$, $M_{r+r_1+k} \geq \nu^{2k} K$ thanks to the definition of r_2, $H_{r+r_1+k}(\nu^k)$ occurs, and, in particular, $M_{r+r_1+k} \geq K$, and $\nu^k \leq M_{r+r_1+k}^{1/2}$. By Lemma 5, on R_1,

$$\mathrm{P}\left[\, D_3(r, r_1, k) \,\middle|\, \mathcal{F}_{r+r_1+k}\right] \geq v(k) \mathbf{1}_{D_3(r,r_1,k-1)},$$

where we put:

$$v(k) = 1 - C_1 \exp\left(-C_2(\nu^{2k} K)^{1/2}\right) - \exp\left(-C_3 \nu^k\right).$$

We conclude that, on R_1,

$$\mathrm{P}\left[\, R_2 \,\middle|\, \mathcal{F}_{r+r_1}\right] \geq \prod_{k=0}^{r_2-1} v(k).$$

Let

$$R_3 = \{\forall n \geq r + r_1 + r_2, f^*(X_{n+1}) = f^*(X_n) + 1, M_n \geq \lambda p_n\}.$$

For $n < r + r_1 + r_2$, we set $D_4(r, r_1, r_2, n) = \Omega$. For $n \geq r + r_1 + r_2$,

$$D_4(r, r_1, r_2, n) = \bigcap_{i=r+r_1+r_2}^{n} D_2(i, (\lambda K^{-1} p_i)^{1/2}).$$

We have

$$D_4(r, r_1, r_2, +\infty) \cap R_2 \subset R_3.$$
$$\mathrm{P}\,[\, D_4(r, r_1, r_2, n) \mid \mathcal{F}_n] = \mathrm{P}\Big[\, D_2(n, (\lambda K^{-1} p_n)^{1/2}) \;\Big|\; \mathcal{F}_n \Big] \mathbf{1}_{D_4(r, r_1, r_2, n-1)}.$$

On $D_4(r, r_1, r_2, n-1) \cap R_2$, $M_n \geq \lambda p_n$, $H_n\left((\lambda K^{-1} p_n)^{1/2}\right)$ occurs, and in particular $M_n \geq K$, and $(\lambda K^{-1} p_n)^{1/2} \leq M_n^{1/2}$, so that, by Lemma 5, on R_2,

$$\mathrm{P}\,[\, D_4(r, r_1, r_2, n) \mid \mathcal{F}_n] \geq w(n) \mathbf{1}_{D_4(r, r_1, r_2, n-1)},$$

where we put:

$$w(n) = 1 - C_1 \exp\left(-C_2 (\lambda p_n)^{1/2}\right) - \exp\left(-C_3 (\lambda K^{-1} p_n)^{1/2}\right).$$

We conclude that, on R_2,

$$\mathrm{P}\left[\, R_3 \mid \mathcal{F}_{r+r_1+r_2} \right] \geq \prod_{i \geq r+r_1+r_2} w(i),$$

the above infinite product being well-defined and positive according to A1. Setting

$$\Xi_1 = \prod_{k \geq 0} v(k)\ ,\ \ \Xi_2 = \prod_{i \geq n_1} w(i),$$

we get that, if r is a record epoch,

$$\mathrm{P}\,[\, R_1 \cap R_2 \cap R_3 \mid \mathcal{F}_r] \geq g^{r_1} \times \Xi_1 \times \Xi_2 > 0.$$

Let

$$R = \left\{ \exists\, r\ ,\ \forall n \geq r \ \ f^*(X_{n+1}) = f^*(X_n) + 1 \right\}$$

We have:

$$R_1 \cap R_2 \cap R_3 \subset R.$$

By Proposition 4, there is, conditional on non-extinction, an infinite number of record epochs. We conclude, thanks to Lemma 13, that conditional on non-extinction, R occurs almost surely. □

Proposition 7. *Conditional on non-extinction, there exists* $\lambda > 0$ *such that, for all* $a \geq 1$, *and all* $n \geq k_a$ *(*k_a *is random),*

$$p_n^{-1} Z_n([0, f^*(X_n) - a]) \leq 4(\theta\lambda)^{-1} \exp(-\beta a).$$

Proof of Proposition 7. We use Proposition 6 and its proof. Conditional on non-extinction, there exists almost surely k such that, for all $n \geq k$, $M_n \geq \lambda p_n$ and $H_n((\lambda K^{-1} p_n)^{1/2})$ occurs. For such an n, $Z_{n+1}([0, f^*(X_{n+1}) - a])$ follows, conditionally to $\mathcal{G}_n$, a binomial law $\mathcal{B}(p_{n+1}, \phi_n)$, where

$$\phi_n \leq \exp(-\beta a) p_n (M'_n)^{-1}.$$

One can easily check that, conditional on $\mathcal{G}_n$, $M_n \geq \lambda p_n$ and $H_n((\lambda K^{-1} p_n)^{1/2})$,

$$\{M'_n \geq (1/2)\theta\lambda p_n, Z_{n+1}([0, f^*(X_{n+1}) - a]) \leq 2\lambda\phi_n\}$$

occurs with probability greater than $1 - C_{14} p_n \exp(-C_{15}(p_n)^{1/2})$. The second Borel-Cantelli lemma (see [7] p.240) implies the result, according to A1. □

4. The random objective function case

We keep the same notations as before. Record epochs are defined exactly as in the Galton-Watson case. Here, $\mathcal{F}_n$ is the σ-algebra generated by X_k, Y_k and $f(\mathbb{T}_{k+1})$, for $0 \leq k \leq n-1$, and X_n. The $\sigma-$algebra $\mathcal{G}_n$ is generated by $\mathcal{F}_n$, $f(\mathbb{T}_{n+1})$ and Y_n. The analog of Proposition 4 is

Proposition 8. *Almost surely on the random objective function f and on the realization of the algorithm, there is an infinite number of record epochs.*

Proof of Proposition 8 (sketch). We have to prove that $f^*(X_n)$ is not bounded from above. The difference from the Galton-Watson case is that f^* may take arbitrarily large negative values, so we have to work a bit more. For a fixed $\lambda > 0$, we define a_n as the largest integer a such that:

$$Z_n([a, f^*(X_n)]) \geq \lambda p_n.$$

We first prove that a_n cannot decrease more than an almost surely finite number of times. Conditionally to $\mathcal{F}_n$, the number of particles that are fixed by mutation is greater than $(1-\epsilon)\mu\lambda p_n$ with probability greater than $1 - \exp(-C(\epsilon)\lambda p_n)$. A result analogous to Lemma 9 shows that, conditionally to $\mathcal{G}_n$ and to the inequality above, $Z_{n+1}([a_n, f^*(X_{n+1})])$ follows a binomial law $\mathcal{B}(p_{n+1}, \phi_n)$ with $\phi_n \geq \zeta\lambda$, where $\zeta > 1$ for ϵ and λ small enough, thanks to A7. Thus, the probability that $a_{n+1} < a_n$ is less than $\exp(-C_1 p_n)$. The Borel-Cantelli lemma proves the stated result. As a consequence, $f^*(X_n)$ is almost surely bounded from below. Thus, if $f^*(X_n)$ is bounded from above with positive probability, two situations may occur. Either there exists $x \in \mathbb{T}$ such that $f(x) = \limsup_n f^*(X_n)$, and such that $x \in X_n$ and $f^*(X_n) = f(x)$ an infinite number of times. In which case an argument similar to the proof in the Galton-Watson case works to get a contradiction. Or there is an infinite number of times n such that X_n contains a newly discovered vertex y in $\mathbb{T}$ such that $f(y) = \limsup_n f^*(X_n)$. Conditional on this, the probability that $f(y(1)) = f(y) + 1$ and that $y(1) \in X_{n+1}$ is bounded from below by a constant. By Lemma 13, this occurs an infinite number of times and we get a contradiction as well. □

Lemma 5 holds as in the Galton-Watson case. A sketch of the proof follows. Conditionally to $\mathcal{F}_n$ and B_n, n is a record epoch and the increments $f(x_i(j)) - f(x_i)$ form a collection of i.i.d. ± 1 random variables. Let

$$J_n = \sum_{i=1}^{L_n} \sum_{j=1}^{m} \mathbf{1}_{\{f(x_i(j)) = f^*(X_n)+1\}} N_n(x_i)$$

and let

$$E_1 = \{J_n \geq (1-\epsilon) m \rho M_n\}.$$

Lemma 10 shows that, on B_n,

$$\mathrm{P}\left[E_1^c \mid \mathcal{F}_n \right] \leq \exp\left(-C_1(\epsilon) M_n / T_n^*\right).$$

Define now

$$E_2 = \{M_n' \geq (1-\epsilon)^2 (1-\mu) \rho m (m+1)^{-1} M_n,\ f^*(Y_n) = f^*(X_n) + 1\}.$$

Lemma 11 shows that, on $B_n \cap E_1$,

$$\mathrm{P}\left[E_2^c \mid \mathcal{F}_n, f(\mathbb{T}_{n+1}) \right] \leq \exp\left(-C_2(\epsilon) J_n\right).$$

The end of the proof is similar to the Galton-Watson case. With the two above results, that is, the existence of an infinite number of record epochs, and Lemma 5, we can prove propositions 6 and 7 exactly as in the Galton-Watson case.

5. Technical lemmas

Lemma 9. *The number of particles at level $f^*(Y_n)$ put by selection in X_{n+1} follows, conditionally to Y_n, a binomial law $\mathcal{B}(p_{n+1}, \mu_{n+1})$. For all $\psi < \exp\beta$, there exists $\lambda_0 > 0$ such that, for all $0 < \lambda < \lambda_0$,*

$$\mu_{n+1} \geq \psi \inf\left(\lambda,\ M_n' p_n^{-1}\right).$$

Proof. By definition of the proportional selection, the number of particles at level $f^*(Y_n)$ put by selection in X_{n+1} follows, conditionally to Y_n, a binomial law $\mathcal{B}(p_{n+1}, \mu_{n+1})$, with

$$\mu_{n+1} = M_n' \exp\left(\beta f^*(Y_n)\right) \left(\sum_{i=1}^{p_n} \exp \beta f\left(Y_n^{(i)}\right) \right)^{-1}.$$

Denote by $\mathcal{H}_n$ the set of indices $1 \leq i \leq p_n$ such that $f\left(Y_n^{(i)}\right) = f^*(Y_n)$. For all $i \notin \mathcal{H}_n$, we have, by definition, $f\left(Y_n^{(i)}\right) \leq f^*(Y_n) - 1$. Thanks to this remark,

$$\sum_{i \notin \mathcal{H}_n} \exp\left(\beta f\left(Y_n^{(i)}\right)\right) \leq \exp(-\beta) \exp\left(\beta f^*(Y_n)\right)(p_n - M_n').$$

Simplifying by $\exp(\beta f^*(Y_n))$ in the expression of μ_{n+1}, we conclude that:

$$\mu_{n+1} \geq M'_n \left(M'_n + (p_n - M'_n)\exp(-\beta)\right)^{-1}.$$

The assertions stated in the lemma is an easy consequence of this inequality. □

Lemma 10. *Let $\chi_1, \ldots, \chi_k$ be k nonnegative real numbers, χ^* be the maximum of the χ_i. Let $s_1, \ldots, s_k$ be i.i.d. random variables with common law supported by $[0, a]$. Let $\theta = \mathrm{E}s_1$. Then, for all $\delta > 0$, all $\gamma \in (0, \delta)$, there exists $\eta > 0$ such that:*

$$\mathrm{P}\left[\sum_{i=1}^{k} s_i\chi_i \leq (1-\delta)\theta\sum_{i=1}^{k}\chi_i\right] \leq \exp\left(-\eta(\delta-\gamma)\theta(\chi^*)^{-1}\sum_{i=0}^{k}\chi_i\right).$$

Proof. By the exponential Chebyshev bound, this probability is bounded from above, for all $t \geq 0$, by:

$$\mathrm{E}\left(\exp\left(-t\sum_{i=1}^{k} s_i\chi_i\right)\right)\exp\left(t(1-\delta)\theta\sum_{i=1}^{k}\chi_i\right).$$

By independence,

$$\mathrm{E}\left(\exp\left(-t\sum_{i=1}^{k} s_i\chi_i\right)\right) = \exp\left(\sum_{i=1}^{k}\log \mathrm{E}(\exp(-ts_i\chi_i))\right).$$

For u near zero, $\mathrm{E}(\exp(-us_1)) = 1 - u\theta + o(u)$. In particular, for all $\gamma > 0$, there exists $\eta > 0$ such that, for all $t \in (0, \eta/\chi^*)$,

$$\log \mathrm{E}(\exp(-ts_i\chi_i)) \leq -(1-\gamma)t\theta\chi_i.$$

Hence an upper bound for the original probability is:

$$\exp\left(-t\theta(\delta-\gamma)\sum_{i=1}^{k}\chi_i\right).$$

□

Lemma 11. *Let $\chi_1, \ldots, \chi_k$ be k nonnegative integers, and $s_1, \ldots, s_k$ in $[0, 1]$. Let $V = \sum_{i=1}^{k} s_i\chi_i$. Consider now $Z_1, \ldots, Z_k$ independent binomial random variables with respective binomial laws $\mathcal{B}(\chi_i, s_i)$. Then, for all $\delta \in (0, 1)$, there exists $C(\delta) > 0$ such that:*

$$\mathrm{P}\left[\sum_{i=1}^{k} Z_i \leq (1-\delta)V\right] \leq \exp(-C(\delta)V).$$

Proof. The Laplace transform is given by:

$$\mathrm{E}\left(\exp\left(-t\sum_{i=1}^{k} Z_i\right)\right) = \exp\left(\sum_{i=1}^{k} \chi_i \log\left(1 + s_i(e^{-t} - 1)\right)\right).$$

For all $0 < \gamma' < 1$, there exists $\phi > 0$ such that, for all $0 \le t \le \phi$,

$$\exp\left(\sum_{i=1}^{k} \chi_i \log\left(1 + s_i(e^{-t} - 1)\right)\right) \le \exp\left(-t(1-\gamma')\sum_{i=1}^{k} s_i \chi_i\right).$$

We conclude by an argument similar to the one used in the previous lemma. □

Lemma 12. *Let* $0 < p < 1$*, and let* S *be a random variable with binomial law* $\mathcal{B}(n, p)$*. For all all* $0 < \epsilon < 1$*, we have:*

$$\mathrm{P}\left[S \le n(1-\epsilon)p\right] \le \exp\left(-(1/2)\epsilon^2 np\right).$$
$$\mathrm{P}\left[S \ge n(1+\epsilon)p\right] \le \exp\left(-(1/2)\epsilon^2 np\right).$$

Proof. We use the exponential Chebyshev bound, valid for every $t > 0$:

$$\mathrm{P}\left[S \le n(1-\epsilon)p\right] \le \exp\left(n(1-\epsilon)pt\right) \times \mathrm{E}\exp\left(-tS\right),$$

with $t = \epsilon$. The inequality:

$$\exp(-\epsilon) - 1 \le \epsilon^2/2 - \epsilon$$

yields the first stated inequality. We proceed likewise to get the second inequality. □

Lemma 13. *Let* $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space with a filtration* $(\mathcal{F}_p)_{p\ge 0}$*, an event* $A \in \sigma(\mathcal{F}_p, p \ge 0)$*, and* $(\tau_i)_{i\ge 1}$ *a family of not necessarily finite stopping times. We assume that for all i and k*

$$\mathrm{P}\left[A \mid \mathcal{F}_k\right]\mathbf{1}_{\{\tau_i = k\}} \ge \gamma \mathbf{1}_{\{\tau_i = k\}},$$

where γ *is a positive constant. Let*

$$B = \{\forall n, \exists i\ ,\ n \le \tau_i < +\infty\}.$$

Then $\mathrm{P}\left[B \setminus A\right] = 0$.

Proof. Define, for $n \ge 0$, $s_n = \inf\{\tau_i : \tau_i \ge n\}$, with the convention $\inf \emptyset = +\infty$. Note that s_n is a $(\mathcal{F}_p)_{p\ge 0}-$ stopping time. On B, each s_n is finite, hence

$$\mathbf{1}_B \le \sum_{u=n}^{+\infty} \mathbf{1}_{\{s_n = u\}} \le 1. \tag{2}$$

For all $u \ge n$, we have:

$$\mathrm{E}\left(\mathbf{1}_{\{s_n=u\}}\mathbf{1}_A \;\middle|\; \mathcal{F}_u\right) \ge \gamma \mathbf{1}_{\{s_n=u\}}.$$

Conditioning by $\mathcal{F}_n$ and summing, (2) implies that:

$$\mathrm{E}\left(\mathbf{1}_A \mid \mathcal{F}_n\right) \geq \gamma \mathrm{E}\left(\mathbf{1}_B \mid \mathcal{F}_n\right).$$

Thanks to the Lévy theorem, see [7] p.263, and to the obvious fact that $B \in \sigma(\mathcal{F}_p, p \geq 0)$, taking the limit $n \to +\infty$ in the above inequality yields:

$$\mathbf{1}_A \geq \gamma \mathbf{1}_B \mathrm{P} - a.s.$$

□

Acknowledgements. We gratefully acknowledge D. Piau for useful comments.

References

1. Aldous, D.: A Metropolis-type optimization algorithm on the infinite tree. Algorithmica **22**, 388–412 (1998)
2. Bérard, J., Bienvenüe, A.: Convergence of a genetic algorithm with finite population. In: D. Gardy, A. Mokkadem (eds.), Mathematics and Computer Science. Algorithms, Trees, Combinatorics and Probabilities. Trends in Mathematics, Basel, 2000. Birkhäuser Verlag, pp. 155–163
3. Bérard, J., Bienvenüe, A.: Un principe d'invariance pour un algorithme génétique en population finie. C. R. Acad. Sci. Paris Sér. I Math. **331**, 469–474 October (2000)
4. Cerf, R.: Asymptotic convergence of genetic algorithms. Adv. Appl. Probab. **30** (2), 521–550 (1998)
5. Del Moral, P., Kallel, L., Rowe, J.: Modelling Genetic Algorithms with Interacting Particle Systems. In: L. Kallel, B. Naudts, A. Rogers (eds.), Natural Computing Series: Theoretical Aspects of Evolutionary Computing, Springer Verlag, Berlin, 2001, pp. 10–67
6. Dembo, A., Gantert, N., Peres, Y., Zeitouni, O.: Large Deviations for Random Walks on Galton-Watson Trees: Averaging and Uncertainty. Probab. Theory Relat. Fields **122** (2), 241–288 (2002)
7. Durrett, R.: Probability: theory and examples. Duxbury Press, Belmont, CA, second edition, 1996
8. Holland, J.H.: Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, Mich., 1975. An introductory analysis with applications to biology, control, and artificial intelligence
9. Hughes, B.: Random walks and Random environments. Vol. **2**. Random environments. The Clarendon Press, Oxford University Press, New York, 1996
10. Lyons, R., Pemantle, R., Peres, Y.: Biased random walks on Galton-Watson trees, Probab. Theory Relat. Fields **106**, 249–264 (1996)
11. Mazza, C., Piau, D.: On the effect of selection in genetic algorithms. Random Structures and Algorithms **18**, 185–200 (2001)
12. Pemantle, R.: Tree-indexed processes. Statist. Sci. **10**, 200–213 (1995)
13. Rabinovich, Y., Wigderson, A.: Techniques for bounding the convergence rate of genetic algorithms. Random Structures and Algorithms **14** (2), 111–138 (1999)
14. Rudolph, G.: Convergence Properties of Evolutionary Algorithms. Hamburg: Kovac, 1997
15. Vose, M.: The Simple Genetic Algorithm. Foundations and theory. MIT Press, Cambridge MA, 1999