Rigorous results on the first hitting times of some Markov chains

J. Garnier

April 13, 1998

Abstract

A comparative study of two mutation operators is proposed. The first is a one bit-flip mutation per individual. The second independently flips each bit with probabilty 1/l.

For each mutation case, expected convergence time is calculated: First, on a needle in a haystack problem, with a (1,1)-ES. Second, on a onemax problem, with a (1+1)-ES.

Introduction

The choice of mutation operator is a crutial issue in all Evolutionary Algorithms (EAs). Even if theoretical convergence results in [4] are proven for any irreductible mutation operators, [1] convergence time can vary significantly according to the considered mutation. Rudolph [8] shows that for large values of ℓ the (1+1)-ES is a lot quicker than the single-bit-flip hill-climber in optimizing long k-path problem with fixed k values (exponential versus polynomial time in ℓ). The difference between the two algorithms is that the (1+1)-ES can do more than one bit-flip at the same time: each bit is flipped with $1/\ell$ probability resulting in 1-bit flip per individual in average.
Rudolph also studies the distribution of the random variable K representing the number of flipped bits in a string of length ℓ:

$$P[K=k] = (C_k^\ell) p^k (1-p)^{\ell-k} \simeq \frac{e^{-1}}{k!}$$

with $p = 1/\ell$.
This directly implies that, for big ℓ values, about one third (e^{-1}) of the individuals are not mutated at all.

This raises a serious question about the efficiency of such a mutation: such a big amount of non-mutated individuals causes a consequent waste of time, and of potential exploration.
In this paper we propose a comparative study of both mutations. Though, comparision result is necessarily problem dependant: Mutation influence is combined with that of fitness-based selection, and we cannot separate their effects.
To avoid such a bias, we consider a mutation operator in the absence of selection. Simply by performing a random mutation walk, with an arbitrary chosen target point. This should

[1] if there exist a path that allows to go from any point of the search space to another via successive mutations

allow to compare *explorative* properties of different mutation operators, in the absence of any landscape interaction.
The second part of this paper, adresses the opposite point of view: that of a maximal landscape interaction. Rather than a needle in a haystack, a onemax problem is studied. The onemax function introduced by [1] has been widely studied in GA litterature:
Bäck [2] derived the transition probabilities for the markov chain with a (1+1)-ES. Muhlenbein [7] considers a simpler markov chain that has (provable) worse performance. He gives an upper bound for convergence time. This bound is minimal with a 1/l mutation, and equals $\ell log(\ell)$. Bäck numerically appoximates the expected absorption time of the transition matrix, and validates $\ell log(\ell)$ bound as a close approximation of the true behavior[3].
Rudolph [9] widely comments these studies as well as the approximations validity. He also prooves an upper bound of $\ell log(\ell)$ for the 1-bit-fip mutation (1+1)-ES.
The influence of crossover, have also been investigated. Culberson [5] shows that the number of local optima in CX landscape increase exponentially in problem size. Nevertheless, to circumvent these local optima, it is essential that crossover takes place between fairly close parents[6].
But in this paper we only focus on (1+1)-ES with mutation only. Exact convergence time for both mutations (1-bit flip and 1/l) is calcultaed. and the relative efficency of both mutations remains unknown also.

1 Search of a point without selection

State space. In this paper l stands for an integer, and we shall develop our theory in the asymptotic framework $l \gg 1$. The state space of the processes that will be studied is $E = \{0,1\}^l$, whose cardinal is 2^l. We shall consider the Hamming distance d defined between two strings $X, X' \in E$ by:

$$d(X, X') = \sum_{i=1}^{l} \mathbb{I}_{X_i \neq X'_i}.$$

Process X. The process $X(n)$ is a process starting from some point X^0, whose evolution is governed by the following way. At each time n, you build the string $X(n)$ by inverting each bit of $X(n-1)$ independently with probability $1/l$. The process $X(n)$ is a Markov chain whose transition matrix of size $2^l \times 2^l$ defined by:

$$P_{X,X'} := \mathbb{P}(X(1) = X'/X(0) = X),$$

is given by:

$$P_{X,X'} = l^{-d(X,X')} \left(1 - l^{-1}\right)^{l-d(X,X')}.$$

Process $\bar{X}$. The process $\bar{X}(n)$ is a process starting from some point X^0, whose evolution is governed by the following way. At each time n, you choose a random integer uniformly over $\{1, ..., l\}$. Then you invert the corresponding bit of the string $\bar{X}(n-1)$. The resulting string is $\bar{X}(n)$. The process $\bar{X}$ is a Markov chain whose transition matrix is:

$$\bar{P}_{X,X'} = \begin{cases} 0 & \text{if } d(X, X') \neq 1 \\ l^{-1} & \text{if } d(X, X') = 1 \end{cases}$$

Goal of the study. We fix some point X^f and we look for the statistical distributions of the respective first hitting times T and $\bar{T}$ of the point X^f by the chains $X(n)$ and $\bar{X}(n)$.

1.1 Main results

PROPOSITION **1.1**
If $d(X^0, X^f) \neq 0$, then:

$$\mathbb{E}[T] \simeq \frac{2^l}{1-e^{-1}}, \qquad \mathbb{E}[\bar{T}] \simeq 2^l,$$

in the sense that $2^{-l} \times \mathbb{E}[T]$ (resp. $2^{-l} \times \mathbb{E}[\bar{T}]$) goes to $1/(1-e^{-1}) \simeq 1.58$ (resp. 1) when $l \to \infty$.

REMARK **1.2**
We can compare these results with the mean values the first hitting times of X^f by some other simple processes. Let us consider the following processes:
1. Random process
At each time you choose some string uniformly over the state space E. The first hitting time obeys the distribution:

$$\mathbb{P}(T_1 = n) = 2^{-l}(1-2^{-l})^{n-1}$$

whose mean is:

$$\mathbb{E}[T_1] = 2^l.$$

2. Exhaustive process
You choose some arbitrary order relation over E and the strings of E are tested one after the other according to the order relation. The first hitting time obeys a uniform distribution over $\{1, ..., 2^l\}$ whose mean is:

$$\mathbb{E}[T_2] = 2^{l-1}.$$

The second method presents the advantage that it never tests twice the same string, and consequently it looses less time than the first one. It has to be noticed that the strategy corresponding to the Markov chain $X(n)$ is the less efficient.

We can be more precise and give the complete statistical distributions of T and $\bar{T}$ in the framework $l \gg 1$.

PROPOSITION **1.3**
If $d(X^0, X^f) \neq 0$, then $2^{-l}(1-e^{-1})T$ (resp. $2^{-l}\bar{T}$) converges in distribution to a random variable with exponential distribution and mean 1 as $l \to \infty$.

This Proposition has two straightforward corollaries.

COROLLARY **1.4**
If $d(X^0, X^f) \neq 0$, then the variances of the variables T and $\bar{T}$ are equal to their respective means as $l \to \infty$.

$$\mathbb{E}\left[(T - \mathbb{E}[T])^2\right] \simeq \frac{2^{2l}}{(1-e^{-1})^2}, \qquad \mathbb{E}\left[(\bar{T} - \mathbb{E}[\bar{T}])^2\right] \simeq 2^{2l}.$$

COROLLARY **1.5**
Assume that we deal with a set of N processes X_i (resp. $\bar{X}_i$) which evolve independently. Then the global first hitting time T (resp $\bar{T}$), which is the infimum of the N independent hitting times, is such that the statistical distribution of $T2^{-l}$ (resp. $\bar{T}2^{-l}$) converges to that of an exponential random variable with mean $N^{-1}(1-e^{-1})^{-1}$ (resp. N^{-1}). In particular its expectation is:

$$\mathbb{E}[T] \simeq \frac{2^l}{N(1-e^{-1})}, \qquad \mathbb{E}[\bar{T}] \simeq \frac{2^l}{N},$$

1.2 Study of the chain $\bar{X}(n)$

Let us introduce the auxiliary Markov process $Z(n)$ with state space $\{0,...,l\}$ starting from $Z(0) = k_f := d(X^0, X^f)$ and with transition matrix $\bar{Q}$:

$$\bar{Q}_{k,k'} = \begin{cases} 0 & \text{if } k' \neq k-1, k+1 \\ 1-\frac{k}{l} & \text{if } k' = k+1 \\ \frac{k}{l} & \text{if } k' = k-1 \end{cases}$$

It is easy to check that:

$$d(\bar{X}(n), X^f) \overset{law}{=} Z(n). \tag{1}$$

Obviously the first hitting time of the point 0 by $Z(n)$ corresponds to the first hitting time $\bar{T}$ of the point X^f by the chain $\bar{X}(n)$.
Let us denote by ν_k the first hitting time of the point k by the chain $Z(n)$ and by $\bar{\nu}_k$ the first hitting time ≥ 1 of the point k. These times coincide if the chain starts from a point distinct from k. But if the chain starts from k, then $\nu_k = 0$ while $\bar{\nu}_k \geq 2$.

LEMMA **1.6**
1. If $k_f = 0,1,2,3$, we have for any l:

$$\mathbb{E}_0[\nu_0] = 0,$$

$$\mathbb{E}_1[\nu_0] = 2^l - 1,$$

$$\mathbb{E}_2[\nu_0] = (2^l - 2)\frac{l}{l-1} + 1,$$

$$\mathbb{E}_3[\nu_0] = (2^l - 2)\frac{l^2 - 2l + 2}{l^2 - 3l + 2} - \frac{4}{l-2},$$

where $\mathbb{E}_k$ stands for the expectation with respect to the statistical distribution of the chain $Z(n)$ starting from k.
2. If $k_f = 1, ..., l$, then we have for $l \gg 1$:

$$\mathbb{E}_{k_f}[\nu_0] \simeq 2^l,$$

which means that $2^{-l} \times \mathbb{E}_{k_f}[\nu_0]$ goes to 1 as $l \to \infty$.

Proof. Obviously $\mathbb{E}_0[\nu_0] = 0$. Let us assume in the following that $k_f \geq 1$. There exists a unique invariant probability measure for the chain $Z(n)$ (i.e. $\mu\bar{Q} = \mu$) which is given by:

$$\mu(k) = 2^{-l} C_l^k.$$

By the general theory of Markov chains, since $Z(n)$ is irreducible and recurrent [3, ch.5, exercice 3.14], the first hitting time $\bar{\nu}_k$ satisfies:

$$\mathbb{E}_k[\bar{\nu}_k] = \frac{1}{\mu(k)}.$$

In particular:

$$\mathbb{E}_0[\bar{\nu}_0] = 2^l. \tag{2}$$

Starting from 0, the first jump is non random: it occurs at time 1 and the process goes to the point 1. Thus:

$$\mathbb{E}_0[\bar{\nu}_0] = \mathbb{E}_0[\bar{\nu}_0/Z(1) = 1] = 1 + \mathbb{E}_0[\bar{\nu}_0 - 1/Z(1) = 1].$$

Applying the weak Markov property (cf Appendix):

$$\mathbb{E}_0[\bar{\nu}_0] = 1 + \mathbb{E}_1[\nu_0],$$

which implies:

$$\mathbb{E}_1[\nu_0] = 2^l - 1. \tag{3}$$

Let $k \in \{0, ..., l-1\}$. Starting from $k+1$, the chain Z jumps either in k or in $k+2$. We can therefore split $\mathbb{E}_{k+1}[\nu_k]$ into two terms:

$$\mathbb{E}_{k+1}[\nu_k] = \mathbb{E}_{k+1}[\nu_k/Z(1) = k]\mathbb{P}_{k+1}(Z(1) = k) + \mathbb{E}_{k+1}[\nu_k/Z(1) = k+2]\mathbb{P}_{k+1}(Z(1) = k+2)$$

which implies by the weak Markov property:

$$\mathbb{E}_{k+1}[\nu_k] = 1 \times \bar{Q}_{k+1,k} + (1 + \mathbb{E}_{k+2}[\nu_k])\, \bar{Q}_{k+1,k+2}. \tag{4}$$

On the other hand, for any $k' \geq 1$ and for any k:

$$\mathbb{E}_{k+k'}[\nu_k] = \sum_{j=0}^{\infty} \mathbb{E}_{k+k'}[\nu_k/\nu_{k+k'-1} = j]\mathbb{P}_{k+k'}(\nu_{k+k'-1} = j).$$

Since $\mathbb{E}_{k+k'}[\nu_k/\nu_{k+k'-1} = j] = j + \mathbb{E}_{k+k'}[\nu_k - \nu_{k+k'-1}/\nu_{k+k'-1} = j]$ and by the strong Markov property:

$$\mathbb{E}_{k+k'}[\nu_k - \nu_{k+k'-1}/\nu_{k+k'-1}] = \mathbb{E}_{k+k'-1}[\nu_k],$$

we get:

$$\mathbb{E}_{k+k'}[\nu_k] = \mathbb{E}_{k+k'}[\nu_{k+k'-1}] + \mathbb{E}_{k+k'-1}[\nu_k]. \tag{5}$$

Substituting this formula with $k' = 2$ into (4), we get the inductive relation:

$$\mathbb{E}_{k+2}[\nu_{k+1}] = (\mathbb{E}_{k+1}[\nu_k] - 1)\, \frac{k+1}{l-(k+1)} - 1. \tag{6}$$

By iterating formula (5) with $k = 0$ and k' decreasing from k_f to 1:

$$\mathbb{E}_{k_f}[\nu_0] = \sum_{k=0}^{k_f - 1} \mathbb{E}_{k+1}[\nu_k].$$

Substituting (3) and (6) into this sum, we get the complete expression of $\mathbb{E}_{k_f}[\nu_0]$, valid for any l. For large l and $k_f \sim \alpha_f l$, $\alpha_f \in (0,1]$, we then find that:

$$\mathbb{E}_{k_f}[\nu_0] = \mathbb{E}_1[\nu_0] + \mathbb{E}_{k_f}[\nu_1], \qquad \mathbb{E}_{k_f}[\nu_1] \simeq 2^l l \int_{a=0}^{\alpha_f} e^{lf(a)} da,$$

where $f(a) = a\ln a - a + \frac{1}{2}a^2$. In particular f is maximal at $a = 0$, $f(0) = 0$, and $f'(a) \sim \ln(a)$. Computing the integral by using Laplace technique, we get:

$$\mathbb{E}_{k_f}[\nu_1] \simeq \frac{2^l}{\ln l}. \tag{7}$$

which yields the result of the Lemma. □

Let us now study the statistical distribution of $2^{-l}\nu_0$.

LEMMA **1.7**
If $k_f \geq 1$, then $2^{-l}\nu_0$ converges in distribution to a random variable with exponential distribution and mean 1 as $l \to \infty$.

Proof.
Step 1. Tightness of $2^{-l}\nu_0$.
By Lemma 1.6:

$$\sup_{l \in \mathbb{N}} \mathbb{E}_{k_f}[2^{-l}\nu_0] < \infty$$

which is sufficient to establish the tightness (and hence the relative compactness) of the sequence of variables $2^{-l}\nu_0$ [2, theorem II-2-2].
Step 2. Convergence in distribution.
Let a, b be two positive reals. Since the chain Z is Markovian, we get from the strong Markov property that:

$$\mathbb{P}_k(\nu_0 \geq [2^l a] + [2^l b] / \nu_0 \geq [2^l a]) = \sum_{j=1}^{l} \nu_k^a(j)\mathbb{P}_j(\nu_0 \geq [2^l b]) \tag{8}$$

where ν_k^a is the distribution of $Z([2^l a])$ starting from k given $\nu_0 \geq [2^l a]$. On the one hand the following obvious estimate holds true for any $j \geq 1$:

$$\mathbb{P}_j(\nu_0 \geq [2^l b]) \leq \mathbb{P}_1(\nu_0 \geq [2^l b]). \tag{9}$$

On the other hand, conditioning with respect to the first time when the chain Z hits 1:

$$\mathbb{P}_j(\nu_0 \geq [2^l b]) = \int_0^\infty \mathbb{P}_1(\nu_0 \geq [2^l b] - t_1) dP_j(t_1),$$

where $P_j(t_1) = \mathbb{P}_j(\nu_1 \leq t_1)$ is the distribution of ν_1 starting from j. If $\delta > 0$, then Markov inequality yileds:

$$\mathbb{P}_j(\nu_1 \geq [2^l \delta]) \leq \frac{[2^l \delta]}{\mathbb{E}_j[\nu_1]}.$$

Combining with Eq. (7) implies that:

$$\mathbb{P}_j(\nu_1 \geq [2^l \delta]) \leq \frac{1}{\delta \ln l}.$$

We then get:

$$\mathbb{P}_j(\nu_0 \geq [2^l b]) \leq \mathbb{P}_1(\nu_0 \geq [2^l b] - [2^l \delta]) + \frac{1}{\delta \ln l}. \tag{10}$$

Substituting (9) and (10) into (8) establishes:

$$\begin{aligned}\liminf_{l\to\infty} \mathbb{P}_1(\nu_0 \geq [2^l b] - [2^l \delta]) &\leq \liminf_{l\to\infty} \mathbb{P}_k(\nu_0 \geq [2^l a] + [2^l b]/\nu_0 \geq [2^l a]) \\ \leq \limsup_{l\to\infty} \mathbb{P}_k(\nu_0 \geq [2^l a] + [2^l b]/\nu_0 \geq [2^l a]) &\leq \limsup_{l\to\infty} \mathbb{P}_1(\nu_0 \geq [2^l b]).\end{aligned} \tag{11}$$

Let ν be a one weak limit of $2^{-l}\nu_0$. Such a limit exists by Step 1, but needs not to be unique, and may depend on the starting point k. Multiplying by 2^{-l} both members inside the $\mathbb{P}(...)$ terms of (11), we obtain:

$$\mathbb{P}_1(\nu \geq b - \delta) \leq \mathbb{P}_k(\nu \geq a + b/\nu \geq a) \leq \mathbb{P}_1(\nu \geq b)$$

In this expression δ is an arbitrary positive real. Letting $\delta \to 0$, we find that the distribution of ν does not depend on k (take $a = 0$, $b > 0$) and satisfies for any $a, b > 0$:

$$\mathbb{P}(\nu \geq b) = \mathbb{P}(\nu \geq a + b/\nu \geq a).$$

Lemma 3.1 then implies that ν is a random variable with exponential distribution. Since its mean is 1, we then get the uniqueness of the weak limit, and consequently the convergence in distribution of $2^{-l}\nu_0$ to this variable. □

1.3 Study of the chain $X(n)$

The approach is very similar to that of the previous section. Let us introduce the process Z:

$$Z(n) = d(X(n), X^f).$$

$(Z(n))_{n\in\mathbb{N}}$ is a Markov chain with transition matrix Q whose coefficients are for $j, j' \in \{0, ..., l\}$:

$$Q_{j,j'} = \sum_{k=0}^{j} \sum_{k'=0}^{l-j} \mathbb{1}_{k'-k=j'-j} C_j^k C_{l-j}^{k'} l^{-k-k'} (1 - l^{-1})^{l-k-k'}.$$

The chain Z is recurrent and its unique invariant probability measure is μ:

$$\mu(j) = 2^{-l} C_l^j.$$

Let us adopt the following notations: ν_k stands for the first hitting time of the point k by the chain $Z(n)$ and $\bar{\nu}_k$ stands for the first hitting time ≥ 1 of the point k. The expectation with respect to the statistical distribution of the chain Z starting from k is denoted by $\mathbb{E}_k$.

LEMMA **1.8**
If $k_f \geq 1$, then:

$$\lim_{l\to\infty} 2^{-l} \mathbb{E}_{k_f}[\nu_0] = \frac{1}{1 - e^{-1}}$$

Proof. Step 1. If $k \leq 7l/8$, then $\mathbb{E}_{k+1}[\nu_k] \leq 2^l (C_l^k)^{-1} 8e$.
Since the unique invariant probability measure of the chain Z is known, we can claim that the expectation of the hitting time $\bar{\nu}_k$ is:

$$\mathbb{E}_k[\bar{\nu}_k] = \frac{1}{\mu(k)} = 2^l (C_l^k)^{-1}.$$

By conditioning with respect to the first jump, we get that this expectation is bounded below by:

$$\mathbb{E}_k[\bar{\nu}_k] \geq \mathbb{P}_k\left(Z(1) = k+1\right)\left(1 + \mathbb{E}_{k+1}[\nu_k]\right). \tag{12}$$

Since the probability to go from k to $k+1$ is bounded below by the probability that one of the $l-k$ bits which coincide with those of X^f is inverted while the other are not:

$$\mathbb{P}_k\left(Z(1) = k+1\right) \geq (1 - l^{-1})^{l-1}(l-k)l^{-1} \geq (1 - kl^{-1})e^{-1} \geq \frac{1}{8}e^{-1}.$$

Substituting into (12) yields the result.
Step 2. If $k \geq 7l/8$, then $\lim_{l \to \infty} 2^{-l} \mathbb{E}_k[\inf_{j < [7l/8]} \nu_j] = 0$.
Starting from $Z(0) = k$, the probability that $Z(1)$ be smaller than k is bounded below by the probability that the $l-k$ bits which coincide with those of X^f are not inverted and at least one of the k bits which do not coincide with those of X^f is inverted. Thus:

$$\begin{aligned} \mathbb{P}_k\left(Z(1) \leq k-1\right) &\geq (1 - l^{-1})^{l-k} \sum_{j=1}^{k} C_k^j l^{-j} (1 - l^{-1})^{k-j} \\ &\geq e^{-1} \sum_{j=1}^{\infty} \frac{(7/8)^j}{j!} \simeq 0.515 \geq 1/2. \end{aligned}$$

As a consequence, the probability that all the $l/8$ first jumps of the chain $Z(n)$ be left jumps is bounded below by:

$$\mathbb{P}_k\left(\text{ sequence of } l/8 \text{ left jumps over } n \in [1, l/8]\ \right) \geq 2^{-l/8},$$

and conversely:

$$\mathbb{P}_k\left(\text{ no sequence of } l/8 \text{ left jumps over } n \in [1, ml/8]\ \right) \leq (1 - 2^{-l/8})^m.$$

Choosing $m = 2^{l/8} 8\alpha$:

$$\mathbb{P}_k\left(\text{ no sequence of } l/8 \text{ left jumps over } n \in [1, l2^{l/8}\alpha]\ \right) \leq e^{-8\alpha},$$

uniformly with respect to $\alpha > 0$ and $k \geq 7l/8$. Besides, whatever the starting point is, the value of the chain Z is smallet than $7l/8$ just after a sequence of $l/8$ left jumps. Thus, for any $k \geq 7l/8$:

$$\mathbb{P}_k\left(\inf_{j=0,\dots,k-1} \nu_j \geq l2^{l/8}\alpha\right) \leq e^{-8\alpha}. \tag{13}$$

For any positive random variable T we have $\int_0^\infty \mathbb{P}(T \geq t)dt = \mathbb{E}[T]$. As a consequence, integrating (13) with respect to α yields:

$$\mathbb{E}_k\left[\inf_{j=0,\dots,k-1} \nu_j\right] \leq 2^{l/8} l/8,$$

and summing over k, we get that for any $k_0 \geq 7l/8$:

$$\mathbb{E}_{k_0}\left[\inf_{j<7l/8} \nu_j\right] \leq 2^{l/8} l^2/64.$$

Step 3. For any $k \geq 2$, $\lim_{l\to\infty} 2^{-l}\mathbb{E}_k[\nu_1] = 0$.
If $k \geq 7l/8$, then:

$$\mathbb{E}_k[\nu_1] \leq \mathbb{E}_{[7l/8]}[\nu_1] + \mathbb{E}_k[\inf_{j<[7l/8]} \nu_j].$$

From Step 2, it is sufficient to prove the statement for $k \leq 7l/8$. If $k \leq [7l/8]$, then:

$$\mathbb{E}_k[\nu_1] \leq \sum_{j=1}^{k-1} \mathbb{E}_{j+1}[\nu_j].$$

By Laplace technique and Step 1, we can show that the sum in the right member is of order $2^l/\ln l$, which establishes the desired result.
Step 4. $\lim_{l\to\infty} 2^{-l} \sum_{j=1}^{\infty} \frac{e^{-1}}{j!} \mathbb{E}_j[\nu_0] = 1.$
Let us denote by p the probability distribution:

$$p(j) = C_l^j l^{-j}(1-l^{-1})^{l-j} \text{ for } j \in \{0, ..., l\}.$$

We still denote by Q the transition matrix of the chain Z. Conditioning with respect to the first jump, we have for any initial probability measure p_0 over $\{0, ..., l\}$:

$$\mathbb{E}_{p_0}[\nu_0] = \sum_{i=1,j=0}^{l} p_0(i) Q_{i,j}(\mathbb{E}_j[\nu_0]+1) = \sum_{j=0}^{l} (p_0 Q)(j)(\mathbb{E}_j[\nu_0]+1) - p_0(0)(\mathbb{E}_p[\nu_0]+1),$$

which also reads as:

$$\mathbb{E}_{p_0}[\nu_0] = (\mathbb{E}_{Qp_0}[\nu_0]+1) - p_0(0)(\mathbb{E}_p[\nu_0]+1). \tag{14}$$

Let p_0 be a probability measure over $\{0, ..., l\}$. Defining inductively $p_{n+1} = p_n Q$, applying (14) inductively with respect to n and summing, we find that for any n:

$$\mathbb{E}_{p_n}[\nu_0] = \mathbb{E}_{p_0}[\nu_0] + n - \sum_{j=0}^{n-1} p_j(0)(\mathbb{E}_p[\nu_0]+1). \tag{15}$$

By the ergodic theorem for Markov chains [3, ch. 6, theorem 2.8] p_n converges to the unique invariant probability measure μ as n goes to infinity. Dividing by n Eq. (15), and letting $n \to \infty$, we get that, for any l:

$$\mu(0)(\mathbb{E}_p[\nu_0]+1) = 1.$$

Since $\mu(0) = 2^{-l}$ and $p(j) \to \frac{e^{-1}}{j!}$ as $l \to \infty$, we get the desired result.
Step 5. For any $j \geq 1$, $\lim_{l\to\infty} 2^{-l}\mathbb{E}_j[\nu_0] = \dfrac{1}{1-e^{-1}}$.
From the obvious estimate

$$\mathbb{E}_1[\nu_0] \leq \mathbb{E}_j[\nu_0] \tag{16}$$

which holds true for any $j \geq 1$, we can deduce by using Step 4 that:

$$\limsup_{l \to \infty} 2^{-l} \mathbb{E}_1[\nu_0] \leq \frac{1}{1-e^{-1}} \tag{17}$$

Applying the inequality

$$\mathbb{E}_j[\nu_0] \leq \mathbb{E}_1[\nu_0] + \mathbb{E}_j[\nu_1] \tag{18}$$

we get from Steps 3 and 4 that:

$$\liminf_{l \to \infty} 2^{-l} \mathbb{E}_1[\nu_0] \geq \frac{1}{1-e^{-1}} \tag{19}$$

Combining (17) and (19) establishes that:

$$\lim_{l \to \infty} 2^{-l} \mathbb{E}_1[\nu_0] = \frac{1}{1-e^{-1}}. \tag{20}$$

Substituting into (16) and (18) and using Step 3, the proof of the Lemma is complete. □

LEMMA **1.9**
If $k_f \geq 1$, then $2^{-l}(1-e^{-1})\nu_0$ converges in distribution to a random variable with exponential distribution and mean 1 as $l \to \infty$.

Proof. The proof of the convergence in distribution of $2^{-l}\nu_0$ is based on the very same arguments as those of the proof of Lemma 1.7. □

1.4 Heuristic comparison between the chains X and $\bar{X}$

The statistical distributions of the chain $\bar{X}(n)$ and of the corresponding stopping time $\bar{T}$ can actually be derived from the statistical distributions of the chain $X(n)$ and of the corresponding stopping time T. However the reader must be warned that this derivation is not as rigorous as the proofs of the previous Lemmas, in the sense that the error terms are not strictly under control, particularly in Lemma 1.10. But the arguments are rather intuitive, and could be made perfectly rigorous with some more work.
Let us consider a sequence of independent and identically distributed random variables $(N_j)_{j \in \mathbb{N}}$ with distribution:

$$\mathbb{P}(N_1 = j) = \frac{e^{-1}}{j!}, \tag{21}$$

N_1 stands for the statistical distribution of the number of inverted bits during a transition of the process X. This distribution is precisely:

$$\mathbb{P}(\text{ } j \text{ bits are inverted }) = C_l^j l^{-j}(1-l^{-1})^{l-j}$$

which converges to the Poisson distribution (21) as $l \to \infty$. Let us denote by S_n the partial sums:

$$S_n = \sum_{j=1}^{n} N_j,$$

LEMMA **1.10**
In the asymptotic framework $l \to \infty$ we have:

$$(X(n))_{n\in\mathbb{N}} \overset{law}{\simeq} (\bar{X}(S_n))_{n\in\mathbb{N}}$$

Proof. Let X^0, X^1 be two strings of the state space E and denote $d := d(X^0, X^1)$. On the one hand ($\bar{P}^j$ is the j-th power of the matrix $\bar{P}$):

$$\begin{aligned}
\mathbb{P}\left(\bar{X}(S_{n+1}) = X^1/\bar{X}(S_n) = X^0\right) &= \sum_{j=0}^{\infty} \mathbb{P}(N_{n+1} = j)\mathbb{P}\left(\bar{X}(j) = X^1/\bar{X}(0) = X^0\right) \\
&= \sum_{j=0}^{\infty} \frac{e^{-1}}{j!}\bar{P}^j_{X^0,X^1} \\
&\overset{l\to\infty}{\simeq} e^{-1}l^{-d},
\end{aligned}$$

because $\bar{P}^d_{X^0,X^1} = d!l^{-d}$ and $\bar{P}^j_{X^0,X^1} = O(l^{-d-1})$ for any $j \neq d$. On the other hand:

$$\begin{aligned}
\mathbb{P}\left(X(n+1) = X^1/X(n) = X^0\right) &= l^{-d}(1-l^{-1})^{l-d} \\
&\overset{l\to\infty}{\simeq} e^{-1}l^{-d}.
\end{aligned}$$

The comparison of these estimates explains the statement of the Lemma. $\square$

By the renewal theory (cf Appendix and (31)), since $\mathbb{E}[N_1] = 1$, the speed rate of S_n is n. But the chain $X(n)$ jumps over some values taken by the chain $\bar{X}(n)$. More exactly, when $\bar{X}(n)$ hits X^f for the first time at time ν_0, then X has some probability p_{miss} to miss X^f, if ν_0 is strictly between two successive times S_n, S_{n+1}. This probability can be computed by the renewal theory (cf Appendix and (32)) in the asymptotic framework $\nu_0 \gg 1$, which holds true when $l \gg 1$:

$$p_{miss} = e^{-1}.$$

If ν_0^j stands for the j-th hitting time of X^f by the chain $\bar{X}(n)$, then $\bar{T}$ is given by the first ν_0^j which corresponds exactly to S_n for some n. The expectation of $\bar{T}$ can be split over the number J which is the first j when ν_0^j coincides with some S_n:

$$\mathbb{E}[\bar{T}] = \sum_{j=0}^{\infty} \mathbb{E}_{k_f}[\nu_0^j/J = j]\mathbb{P}(J = j).$$

By the strong Markov property, the excursions of the chain $\bar{X}(n)$ between two successive hitting times ν_0^j and ν_0^{j+1} are independent and identically distributed. Thus:

$$\mathbb{E}[\bar{T}] = \sum_{j=0}^{\infty} \left(\mathbb{E}_{k_f}[\nu_0] + \mathbb{E}_0[\bar{\nu}_0]j\right)\mathbb{P}(J = j).$$

Since $\mathbb{P}(J = j) = p_{miss}^{j-1}(1 - p_{miss})$ and $\mathbb{E}_0[\bar{\nu}_0] = 2^l$, we get that:

$$\mathbb{E}[\bar{T}] = \mathbb{E}_{k_f}[\nu_0] + \frac{2^l}{e-1} = \mathbb{E}[T] + \frac{2^l}{e-1}.$$

2 Search of a point with selection

2.1 Process $\bar{X}$

Let us consider the process $\bar{X}$ starting from X^0 whose evolution is governed by the following way. At each time n, you choose a random integer uniformly over $\{1, ..., l\}$. Then you invert the corresponding bit of the string $\bar{X}(n-1)$. The resulting string is denoted by $\tilde{X}(n-1)$. If $d(\tilde{X}(n-1), X^f) < d(\bar{X}(n-1), X^f)$, then $\bar{X}(n) = \tilde{X}(n-1)$, otherwise $\bar{X}(n) = \bar{X}(n-1)$. The transition matrix of the chain $\bar{X}(n)$ is consequently:

$$\bar{P}_{X,X'} = \begin{cases} 1 - d(X, X^f) l^{-1} & \text{if } d(X,X) = 0 \\ l^{-1} & \text{if } d(X,X') = 1 \text{ and } d(X', X^f) = d(X, X^f) - 1 \\ 0 & \text{otherwise} \end{cases}$$

PROPOSITION **2.1**
Assuming that the starting point X^0 is chosen uniformly over the state space $E = \{0,1\}^l$, the normalized first hitting time $T/(l \ln l)$ converges in probability to 1 as $l \to \infty$.
For $l \gg 1$, the expectation of T can be expanded as:

$$\mathbb{E}[T] \overset{l \gg 1}{\simeq} l(\ln l + C - \ln 2)$$

up to terms of order $o(l)$. C is the Euler's constant whose value is $C \simeq 0.58$.

Proof. Let us introduce the auxiliary process $Z(n)$ with state space $\{0, ..., l\}$ defined by:

$$Z(n) = d(\bar{X}(n), X^f). \tag{22}$$

Z is a Markov chain starting from $Z(0) = d(X^0, X^f)$ and whose jumps are governed by the transition matrix $\bar{Q}$:

$$\bar{Q} = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \frac{1}{l} & 1 - \frac{1}{l} & 0 & \dots & \dots & 0 \\ 0 & \frac{2}{l} & 1 - \frac{2}{l} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 - \frac{1}{l} & \frac{1}{l} & 0 \\ 0 & \dots & \dots & 0 & 1 & 0 \end{pmatrix}$$

Obviously the first hitting time of the point 0 by $Z(n)$ corresponds to the first hitting time $\bar{T}$ of the point X^f by the chain $\bar{X}(n)$. Starting from k, the first hitting time of the point 0 by the chain Z is:

$$\bar{T} = \sum_{j=1}^{k} \tau_j,$$

where the τ_j's are the times necessary for the chain to jump from j to $j-1$. By the strong Markov property, the τ_j's are independent random variables with respective distributions ($Q_{j,j}^n$ is the n-th power of the coefficient $Q_{j,j}$):

$$\mathbb{P}(\tau_j = t) = (1 - Q_{j,j} Q_{j,j}{}^n = \frac{j}{l} \left(1 - \frac{j}{l}\right)^{t-1} \quad \text{for } t \in \mathbb{N}^*.$$

In particular their first and second order moments are:

$$\mathbb{E}[\tau_j] = \frac{l}{j}, \qquad \mathbb{E}[\tau_j^2] = \frac{2l^2 - jl}{j^2}$$

We can then deduce the expectation of $\bar{T}$:

$$\mathbb{E}_k[\bar{T}] = \sum_{j=1}^{k} \mathbb{E}[\tau_j] = l \sum_{j=1}^{k} j^{-1} \stackrel{k \gg 1}{\simeq} l(\ln k + C),$$

up to terms of order $o(l)$. If the chain $\bar{X}$ starts from a point X^0 which is chosen uniformly over the state space $\{0,1\}$, then the chain Z starts from the point k_0 which obeys the distribution μ:

$$\mu(k_0) = 2^{-l} C_l^{k_0}.$$

and then:

$$\mathbb{E}_\mu[\bar{T}] = \sum_{k=0}^{l} \mu(k) \mathbb{E}_k[\bar{T}].$$

Using Stirling's formula:

$$\mathbb{E}_\mu[\bar{T}] \stackrel{l \gg 1}{\simeq} l \left(\ln l + C - \sqrt{l} \int_0^1 e^{-lf(a)} \frac{\ln(a)}{\sqrt{2\pi a(1-a)}} da \right),$$

where $f(a) = \ln 2 + a\ln(a) + (1-a)\ln(1-a)$. Since f is minimal at $a = 1/2$, $f(1/2) = 0$, $f'(1/2) = 0$ and $f''(1/2) = 4$, we get by using Laplace technique that:

$$\int_0^1 e^{-lf(a)} \frac{\ln(a)}{\sqrt{2\pi a(1-a)}} da \simeq -\sqrt{\frac{2}{\pi}} \ln 2 \int_{-\infty}^{\infty} e^{-2a^2 l} da \simeq -\frac{\ln 2}{\sqrt{l}}.$$

Consequently:

$$\mathbb{E}_\mu[\bar{T}] \stackrel{l \gg 1}{\simeq} l\,(\ln l + C - \ln 2)$$

up to terms of order $o(l)$. The calculation of the second order moment of $\bar{T}$ can be carried out by the same way:

$$\begin{aligned} \mathbb{E}_k[\bar{T}^2] &= \sum_{j=1}^{k} \mathbb{E}[\tau_j^2] + \sum_{i \neq j=1}^{k} \mathbb{E}[\tau_i]\mathbb{E}[\tau_j] = l^2 \left(\sum_{j=1}^{k} j^{-2} \right) - l \left(\sum_{j=1}^{k} j^{-1} \right) + l^2 \left(\sum_{j=1}^{k} j^{-1} \right)^2 \\ &\stackrel{l \gg 1}{\simeq} l^2 \left(\frac{\pi^2}{6} + C^2 + (\ln k)^2 + 2C \ln k \right) - l(\ln k + C). \end{aligned}$$

So:

$$\mathbb{E}_\mu[\bar{T}^2] \stackrel{l \gg 1}{\simeq} l^2 \left\{ (\ln l)^2 + (2C - 2\ln 2)\ln l + \left(\frac{\pi^2}{6} + C^2 + (\ln 2)^2 - 2C \ln 2 \right) \right\}$$

up to terms of order $o(l^2)$. As a consequence the normalized variance defined by:

$$\bar{V} := \frac{\sqrt{\mathbb{E}[\bar{T}^2] - \mathbb{E}[\bar{T}]^2}}{\mathbb{E}[\bar{T}]}$$

satisfies:

$$\bar{V} \stackrel{l \gg 1}{\simeq} \frac{\pi}{\sqrt{6}(\ln l + C - \ln 2)} \stackrel{l \to \infty}{\longrightarrow} 0$$

at rate $(\ln l)^{-1}$, which implies that $\bar{T}/(l\ln l)$ converges in probability to 1. However the convergence rate of the normalized variance towards 0 is about $(\ln l)^{-1}$, which is slow for practical purposes. The normalized variance is actually of order 1 for a large band of values of l. In order to get $\bar{V} < 10\%$, it is necessary that $l > 4.10^5$.

2.2 Process X

Let us consider the process X starting from X^0 whose evolution is governed by the following way. At each time n, you build a string $\tilde{X}(n-1)$ by inverting each bit of $X(n-1)$ independently with probability $1/l$. If $d(\tilde{X}(n-1), X^f) < d(X(n-1), X^f)$, then $X(n) = \tilde{X}(n-1)$, otherwise $X(n) = X(n-1)$. The transition matrix of the chain $X(n)$ is consequently:

$$P_{X,X'} = \begin{cases} P_{X,X} & \text{if } d(X,X') = 0 \\ l^{-d(X,X')}\left(1-l^{-1}\right)^{l-d(X,X')} & \text{if } d(X',X^f) < d(X,X^f) \\ 0 & \text{otherwise} \end{cases}$$

with $P_{X,X} = 1 - \sum_{k=0}^{d(X,X^f)} \sum_{k'=0}^{l-d(X,X^f)} \mathbb{1}_{k<k'} C^k_{d(X,X^f)} C^{k'}_{l-d(X,X^f)} l^{-k-k'} \left(1-l^{-1}\right)^{l-k-k'}$.

PROPOSITION **2.2**
Assuming that the starting point X^0 is chosen uniformly over the state space $\{0,1\}^l$, the normalized first hitting time $T/(l\ln l)$ converges in probability to e as $l \to \infty$.
For $l \gg 1$, the expectation of T can be expanded as:

$$\mathbb{E}[T] \overset{l\gg 1}{\simeq} l(e\ln l + eC - e\ln 2 + \chi) \ \textit{where}\ \chi \simeq -1.59 \ \textit{and}\ C \simeq 0.58,$$

up to terms of order $o(l)$.

Proof. The strategy of the proof is similar to that of Proposition 2.1. However the details are a little more complicated because the transition matrix of the chain $X(n)$ is more complicated than the one of the chain $\bar{X}(n)$. Let us introduce the auxiliary process $Z(n)$ with state space $\{0,...,l\}$ defined by:

$$Z(n) = d(X(n), X^f). \tag{23}$$

Z is a Markov chain starting from $Z(0) = d(X^0, X^f)$. The coefficients of the corresponding transition matrix Q are given by:

$$Q_{j,j-j'} = \begin{cases} \sum_{k=0}^{(l-j)\wedge(j-j')} C_j^{k+j'} C^k_{l-j} l^{-2k-j'}(1-l^{-1})^{l-2k-j'} & \text{if } j' \in \{1,...,j\} \\ Q_{j,j} & \text{if } j' = 0 \\ 0 & \text{if } j' \notin \{0,...,j\} \end{cases} \tag{24}$$

with $Q_{j,j} = 1 - \sum_{j'=1}^{j} Q_{j,j-j'}$.
Let us denote by τ_j the first jump time when $Z(0) = j$, and by N_j the length of this jump. By the strong Markov property, the random variables τ_j and N_j are independent and their

respective statistical distributions are given by:

$$\begin{aligned} \mathbb{P}(\tau_j = t) &= (1 - Q_{j,j}) Q_{j,j}^{t-1} \text{ for } t \in \mathbb{N}^*, &(25)\\ \mathbb{P}(N_j = n) &= \frac{Q_{j,j-n'}}{1 - Q_{j,j}} \text{ for } n \in \mathbb{N}^*. &(26) \end{aligned}$$

The statistical distribution of the first hitting time T can be expressed in terms of the variables N_j and τ_j. In particular, conditioning with respect to the first jump yields:

$$\mathbb{E}_j[T] = \sum_{t=1}^{\infty} \sum_{n=1}^{j} \mathbb{E}_j[T/\tau_j = t, N_j = n] \mathbb{P}(\tau_j = t) \mathbb{P}(N_j = n)$$

Since the strong Markov property insures that:

$$\mathbb{E}_j[T/\tau_j = t, N_j = n] = t + \mathbb{E}_{j-n}[T],$$

we get the inductive relation:

$$\mathbb{E}_j[T] = \mathbb{E}[\tau_j] + \sum_{n=1}^{j} \mathbb{E}_{j-n}[T] \mathbb{P}(N_j = n) \tag{27}$$

Indeed, iterating this formula establishes the following closed-form expression which holds true for any j and l:

$$\mathbb{E}_j[T] = \sum_{k=1}^{\infty} \sum_{n_1 + \ldots + n_k = j} \left(\sum_{i=0}^{k-1} \mathbb{E}[\tau_{j-n_1-\ldots-n_i}] \right) \prod_{j=1}^{k} \mathbb{P}(N_j = n_j). \tag{28}$$

In the asymptotic framework $l \gg 1$, this complicated expression gets much simpler.
Step 1. There exists functions f_τ and f_N which do not depend on l such that, for $l \gg 1$: $\mathbb{E}[\tau_j] = f_\tau(j/l)$ and $\mathbb{E}[N_j] = f_N(j/l)$.

$$\begin{aligned} \frac{1}{f_\tau(a)} &= 1 - e^{-1} \sum_{k=0}^{\infty} \frac{(1-a)^k}{k!} \left(\sum_{j=0}^{k} \frac{a^j}{j!} \right), &(29)\\ f_N(a) &= \frac{(2a-1)e}{f_\tau(a)} + (1-a)a \left(\sum_{k=0}^{\infty} \frac{(1-a)^k a^k}{k!(k+1)!} \right) + a \left(\sum_{k=0}^{\infty} \frac{(1-a)^k a^k}{k!^2} \right) &(30) \end{aligned}$$

This step results from the study of the definitions (24)-(26) of the statistical distributions of N_j and τ_j.

Step 2. If $a \in (0,1]$, then we have for $l \gg 1$: $\mathbb{E}_{[al]}[T] = \sum_{j=1}^{[al]} \frac{f_\tau(j/l)}{f_N(j/l)}$.
Let $0 < \delta \ll l$, such that $1 \ll \delta l \ll 1$. By Step 1, we get that over the interval $j \in \{[al], ..., [al] - [\delta l]\}$, the statistical distributions of N_j and τ_j are almost identical to the ones of $N_{[al]}$ and $\tau_{[al]}$. Using the inductive relation (27):

$$\mathbb{E}_{[al]}[T] \simeq \sum_{k=1}^{\infty} \sum_{n_1+\ldots+n_{k-1} \leq [\delta l] < n_1 + \ldots + n_k} \prod_{j=1}^{k} \mathbb{P}(N_{[al]} = n_j) \left(k\mathbb{E}[\tau_{[al]}] + \mathbb{E}_{[al]-n_1+\ldots+n_k}[T] \right)$$

FIG 1 : *Functions* $a \mapsto \frac{1}{f_\tau(a)}$ *and* $a \mapsto f_N(a)$.

By the renewal theory (see Appendix and (31)):

$$\frac{1}{[\delta l]}\sum_{k=1}^{\infty} k\mathbb{P}\left(N_1+...+N_{k-1}\leq [\delta l] < N_1+...+N_k\right) \overset{l\to\infty}{\simeq} \frac{1}{\mathbb{E}[N_{[al]}]} = \frac{1}{f_N(a)},$$

which implies that:

$$\mathbb{E}_{[al]}[T] \simeq \frac{f_\tau(a)}{f_N(a)}[\delta l] + \mathbb{E}_{[al]-[\delta l]}[T].$$

Iterating this process completes the proof of Step 2.
Step 3. For $l \gg 1$: $\mathbb{E}_\mu[T] = l(e\ln l + C - \ln 2 + \chi)$ *where* $\chi \simeq -1.59$.
The function $a \mapsto \frac{f_\tau(a)}{f_N(a)}$ has a singularity near the point $a=0$ of the type a^{-1}. Nevertheless $g := a \mapsto \frac{f_\tau(a)}{f_N(a)} - \frac{e}{a}$ is a smooth function. As a consequence:

$$\mathbb{E}_{[al]}[T] = \sum_{j=1}^{[al]}\frac{el}{j} + \sum_{j=1}^{[al]} g(j/l)$$

which yields:

$$\mathbb{E}_{[al]}[T] = l\left(e\ln l - e\ln a + eC + G(a)\right)$$

where G is the integral of g:

$$G(a) = \int_0^a g(s)ds.$$

FIG 2 : *Functions $a \mapsto g(a)$ and $a \mapsto G(a)$.*

The result then follows readily, since the initial measure μ gets sharply peaked close to $l/2$ and $G(1/2) = \chi$.
Step 4. Conclusion.
Adopting the same strategy as in the proof of Proposition 2.1, long but straightforward calculations establish that:

$$\frac{\sqrt{\mathbb{E}[T^2] - \mathbb{E}[T]^2}}{\mathbb{E}[T]} \stackrel{l \to \infty}{\longrightarrow} 0$$

at rate $(\ln l)^{-1}$, which implies that $T/(l \ln l)$ converges in probability to e. □

3 Appendix: Some useful results of probability theory

Markov property

A Markov chain $Z(n)$ satisfies the so-called weak and strong Markov properties [3].
Weak Markov property
For any integer n_0 and for any function ϕ which depends on the whole path:

$$\mathbb{E}\left[\phi\left((Z(n))_{n \geq n_0}\right) / (Z(n))_{n \leq n_0}\right] = \mathbb{E}\left[\phi\left((Z(n))_{n \geq n_0}\right) / Z(n_0)\right].$$

Besides, if the chain is homogeneous (i.e. its transition matrix is time-independent), then:

$$\mathbb{E}\left[\phi\left((Z(n))_{n \geq n_0}\right) / Z(n_0)\right] = \mathbb{E}_{Z(n_0)}\left[\phi\left((Z(n))_{n \geq 0}\right)\right].$$

Strong Markov property
For any stopping time τ and for any function ϕ which depends on the whole path:

$$\mathbb{E}\left[\phi\left((Z(n))_{n\geq\tau}\right)/(Z(n))_{n\leq\tau}\right] = \mathbb{E}\left[\phi\left((Z(n))_{n\geq\tau}\right)/Z(\tau)\right].$$

Besides, if the chain is homogeneous, then:

$$\mathbb{E}\left[\phi\left((Z(n))_{n\geq\tau}\right)/Z(\tau)\right] = \mathbb{E}_{Z(\tau)}\left[\phi\left((Z(n))_{n\geq 0}\right)\right].$$

If τ is a stopping time of the form $\tau = \inf\{n, Z(n) = k\}$, then:

$$\mathbb{E}_{Z(\tau)}\left[\phi\left((Z(n))_{n\geq 0}\right)\right] = \mathbb{E}_k\left[\phi\left((Z(n))_{n\geq 0}\right)\right].$$

Renewal theory

Let $(N_j)_{j\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with distribution $\mathbb{P}(N_1 = k) = p_k$, $k \in \mathbb{N}$.

If we denote by S_n the partial sums $S_n = \sum_{j=1}^{n} N_j$, then for any k, there exists a unique n_k such that $S_{n-1} \leq k < S_n$. Of course the sequence n_k is random, and it satisfies [1]:

$$\frac{\mathbb{E}[n_k]}{k} \overset{k\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}[N_1]}, \tag{31}$$

$$\mathbb{P}\left(\exists n, S_n = k\right) \overset{k\to\infty}{\longrightarrow} \frac{1-p_0}{\mathbb{E}[N_1]}. \tag{32}$$

A lemma about exponential random variables

LEMMA **3.1**
Let T be a $\mathbb{R}^+$-valued random variable. T obeys an exponential distribution if and only if for every $a, b \geq 0$:

$$\mathbb{P}(T \geq a + b/T \geq a) = \mathbb{P}(T \geq b). \tag{33}$$

Proof. If T obeys an exponential distribution, then it obviously satisfies (33). Let us now consider a $\mathbb{R}^+$-valued random variable satisfying (33). We denote:

$$\alpha := \mathbb{P}(T \geq 1)$$

Step 1. For any $n \in \mathbb{N}$, $\mathbb{P}(T \geq n) = \alpha^n$.
Applying (33) with $b = 1$ and $a = n$ yields:

$$\mathbb{P}(T \geq n+1) = \mathbb{P}(T \geq n+1/T \geq n)\mathbb{P}(T \geq n) = \alpha\mathbb{P}(T \geq n).$$

By induction with respect to n we get the result.
Step 2. For any $q \in \mathbb{Q}$, $\mathbb{P}(T \geq q) = \alpha^q$.
Let $q \in \mathbb{Q}$. q is of the form $q = r/n$ where r and n are integers. Applying (33) with $b = q$ and $a = mq$ yields:

$$\mathbb{P}(T \geq q(m+1)) = \mathbb{P}(T \geq q)\mathbb{P}(T \geq mq).$$

By induction with respect to $m = n-1, ..., 0$, we get

$$\mathbb{P}(T \geq qn) = \mathbb{P}(T \geq q)^n.$$

Since $qn = r$ is an integer we have $\mathbb{P}(T \geq qn) = \alpha^r$, which establishes the result.
Step 3. For any $t \in \mathbb{R}^+$, $\mathbb{P}(T \geq t) = \alpha^t$.
The function $t \mapsto \mathbb{P}(T \geq t)$ is monotone over $\mathbb{R}^+$. Step 2 then obviously implies the result of Step 3. □

References

[1] D. H. Ackley. *A connectionnist Machine for Genetic Hill-climbing.* Kluwer, Boston, 1987.

[2] T. Bäck. The interaction of mutation rate, selection, and self-adaptation in genetic algorithm. pages 85–94. Morgan Kaufmann, 1992.

[3] T. Bäck. Optimal Mutation Rate in Genetic Search. In S. Forrest, editor, *Proceedings of the 5^{th} International Conference on Genetic Algorithms*, pages 2–8. Morgan Kaufmann, 1993.

[4] R. Cerf. *Une théorie assymptotique des algorithmes génétiques.* PhD thesis, Université de Montpellier II, March 1994.

[5] J. Culberson and Jonathan Lichtner. On searching 'alpha'-ary hypercubes and related graphs. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms 4*, pages 263–290. Morgan Kaufmann, 1996.

[6] J. Horn and C.R Reeves. The genetic algorithm landscape for the onemax problem. 1996. Submitted to Second Nordic Workshop On Genetic Algorithms.

[7] H. Mühlenbein. How genetic algorithms really work: I. mutation and hill-climbing. In R. Manner and B. Manderick, editors, *Proceedings of the 2^{nd} Conference on Parallel Problems Solving from Nature*, pages 15–25. Morgan Kaufmann, 1992.

[8] G. Rudolph. Convergence of non-elitist strategies. In Z. Michalewicz, J. D. Schaffer, H.-P. Schwefel, D. B. Fogel, and H. Kitano, editors, *Proceedings of the First IEEE International Conference on Evolutionary Computation*, pages 63–66. IEEE Press, 1994.

[9] G. Rudolph. *Convergence Properties of Evolutionary Algorithms.* Kovac, Hamburg, 1997.

References

[1] I. Guikhman and A. Skorokhod, *Introduction à la théorie des processus aléatoires*, Mir, Moscou, 986.

[2] S. Ethier and T. Kurtz, *Markov processes*, Wiley, New York, 1986.

[3] D. Revuz, *Markov chains*, North-Holland, Amsterdam, 1975.