The quasispecies regime for the simple genetic algorithm with roulette–wheel selection

Raphaël Cerf
DMA, École Normale Supérieure
E-mail: rcerf@math.u-psud.fr
June 30, 2015

Abstract
We introduce a new parameter to discuss the behavior of a genetic algorithm. This parameter is the mean number of exact copies of the best fit chromosomes from one generation to the next. We argue that the genetic algorithm should operate efficiently when this parameter is slightly larger than 1. We consider the case of the simple genetic algorithm with the roulette–wheel selection mechanism. We denote by $\ell$ the length of the chromosomes, by $m$ the population size, by $p_C$ the crossover probability and by $p_M$ the mutation probability. We start the genetic algorithm with an initial population whose maximal fitness is equal to $f_0^*$ and whose mean fitness is equal to $\overline{f}_0$. We show that, in the limit of large populations, the dynamics of the genetic algorithm depends in a critical way on the parameter $\pi = \left(\frac{f_0^*}{\overline{f}_0}\right)(1-p_C)(1-p_M)^\ell$. If $\pi < 1$, then the genetic algorithm might operate in a disordered regime: there exist positive constants $\beta$ and $\kappa$ which do not depend on $m$ such that, for some fitness landscapes and some initial populations, with probability larger than $1 - 1/m^\beta$, before generation $\kappa \ln m$, the best fit individual will disappear, and until generation $\kappa \ln m$, the mean fitness will stagnate. If $\pi > 1$, then the genetic algorithm operates in a quasispecies regime: there exist positive constants $\kappa, p^*$ which do not depend on $m$ such that, for any fitness landscape and any initial population, with probability larger than $p^*$, until generation $\kappa \ln m$, the maximal fitness will not decrease and before generation $\kappa \ln m$, the mean fitness will increase by a factor $\sqrt{\pi}$. These results suggest that the mutation and crossover probabilities should be tuned so that, at each generation, maximal fitness $\times (1-p_C)(1-p_M)^\ell >$ mean fitness.
1 Introduction

A central problem to implement efficiently a genetic algorithm is the adjustment of the many parameters controlling the algorithm. If we focus on the classical simple genetic algorithm, these parameters are: the population size, the probabilities of crossover and mutation. There exists a huge literature discussing this question. The main message given by the numerous works conducted over the years is that, contrary to the initial hopes, there exists no universal choice of parameters and the optimal choices depend heavily on the fitness landscape. We refer the reader to [7] for a recent review.

Our goal here is to attract the attention on a single parameter, which somehow sums up the effects of the various mechanisms at work in a genetic algorithm, and which is quite natural from the probabilistic viewpoint. The parameter we have in mind is the mean number of exact copies of the best fit chromosomes from one generation to the next. Let us call it $\pi$. We suggest that, at any generation, the various operators of the genetic algorithm should be controlled in order to ensure that $\pi$ is slightly larger than 1. Indeed, if $\pi < 1$, then the best fit chromosomes are doomed to disappear quickly from the population. If $\pi > 1$, then, with positive probability, the best fit chromosomes will perpetuate and one of them will quickly become the most recent common ancestor of the whole population. It is not desirable that $\pi$ is much larger than 1, in order to avoid the premature convergence of the algorithm. The best situation is when the population retains the best fit individuals and explores actively their neighborhoods. Ideally we would like to have a few copies of the best fit individuals and a cloud of mutants descending from them. This is why we aim at tuning the parameters so that $\pi$ is only slightly larger than 1. An interesting tentative to induce this behavior is what has been called "elitism" in the genetic algorithm literature. Under elitism, the best fit chromosomes are automatically retained from one generation to the next. However, we believe that the resulting dynamics is intrinsically different from the one we are aiming at when tuning the parameters so that $\pi > 1$. Indeed, we wish to build a probabilistic dynamics which automatically focuses the search around the best fit individuals, and it might be that, even using elitism, the best fit individuals are quickly forgotten during the search and none of them has a chance to become the most recent common ancestor.

An advantage with the parameter $\pi$ is that we can easily compute simple bounds in terms of the parameters of the algorithm. This becomes particularly true if we perform in addition an asymptotic expansion in one or several parameters. If we do so, we can even prove rigorous results which strongly support the previous ideas. More precisely, we will consider here the case of large populations. This kind of analysis has been previ-
ously conducted for the simple genetic algorithm with ranking selection [1].
We try here to extend this analysis to the simple genetic algorithm with
roulette–wheel selection. This task turned out to be very difficult, because
the dynamics is very sensitive to the variations of the fitness values. Most
of the results obtained for ranking selection do not hold with roulette–wheel
selection. We present only two results, which demonstrate that, depending
on the parameters and the fitness distribution of the current population,
the genetic algorithm can operate either in a disordered regime, where the
best fit individuals are typically lost, or in a quasispecies regime, where the
best fit individuals survive and invade a positive fraction of the population.
Our results have their roots in the quasispecies theory developed by
Eigen, McCaskill and Schuster [3]. We refer the reader to the introduction
of [1] for a quick summary of the development of these ideas, as well as
for pointers to the numerous relevant references in the genetic algorithm
literature.

We study here the classical simple genetic algorithm with the roulette–
wheel selection mechanism, as described in the famous books of Holland [6]
and Goldberg [4]. We denote by \( \ell \) the length of the chromosomes, by \( m \)
the population size, by \( p_C \) the crossover probability and by \( p_M \) the mutation
probability. We start the genetic algorithm with an initial population whose
maximal fitness is equal to \( f_0^* \) and whose mean fitness is equal to \( f_0 \). We
show that, in the limit of large populations, the dynamics of the genetic
algorithm depends in a critical way on the parameter
\[
\pi = \left( \frac{f_0^*}{f_0} \right) (1 - p_C)(1 - p_M)^\ell.
\]
- If \( \pi < 1 \), then the genetic algorithm might operate in a disordered regime:
  there exist positive constants \( \beta \) and \( \kappa \) which do not depend on \( m \) such that,
  for some fitness landscapes and some initial populations, with probability
  larger than \( 1 - 1/m^3 \), before generation \( \kappa \ln m \), the best fit individual will
disappear and until generation \( \kappa \ln m \), the mean fitness will stagnate.
- If \( \pi > 1 \), then the genetic algorithm operates in a quasispecies regime:
  there exist positive constants \( \kappa, p^* \) which do not depend on \( m \) such that,
  for any fitness landscape and any initial population, with probability greater
  than \( p^* \), until generation \( \kappa \ln m \), the maximal fitness will not decrease and
  before generation \( \kappa \ln m \), the mean fitness will increase by a factor \( \sqrt{\pi} \).

These results suggest that at each generation, the mutation and crossover
probabilities should be tuned so that
\[
\text{maximal fitness} \times (1 - p_C)(1 - p_M)^\ell > \text{mean fitness}.
\]
It seems therefore judicious to choose “large” values of \( p_M \) and \( p_C \) com-
patible with the condition \( \pi > 1 \). This means that the mutation probability
should be of order \( 1/\ell \); more precisely, the condition \( \pi > 1 \) implies that
\[
\ell p_M + p_C < \ln \left( \frac{f_0^*}{f_0} \right).
\]
2 The model

In this section, we provide a brief description of the simple genetic algorithm. The goal of the simple genetic algorithm is to find the global maxima of a fitness function \( f \) defined on \( \{0, 1\}^\ell \) with values in \([0, +\infty[\). We consider the most classical and simple version of the genetic algorithm, as described in Goldberg’s book [4]. The genetic algorithm works with a population of \( m \) points of \( \{0, 1\}^\ell \), called the chromosomes, and it repeats the following fundamental cycle in order to build the generation \( n + 1 \) from the generation \( n \):

Repeat

• Select two chromosomes from the generation \( n \)
• Perform the crossover
• Perform the mutation
• Put the two resulting chromosomes in generation \( n + 1 \)

Until there are \( m \) chromosomes in generation \( n + 1 \)

We use roulette-wheel selection. We use the standard single point crossover and the crossover probability is denoted by \( p_C \). We use independent parallel mutation at each bit and the mutation probability is denoted by \( p_M \). So we use the classical genetic operators and we refer to [1] for their precise descriptions. The only difference is the choice for the selection function. Here we choose the roulette-wheel selection. The probability of selecting the \( i \)-th chromosome \( x(i) \) in the population \( x \) is given by the selection distribution defined by

\[
\text{sel}(x, i) = \frac{f(x(i))}{f(x(1)) + \cdots + f(x(m))}.
\]

3 The results

We start the genetic algorithm with an initial population whose maximal fitness is equal to \( f_0^* \) and whose mean fitness is equal to \( \bar{f}_0 \). We denote by \( x_0 \) the initial population and by \( x_0(1), \ldots, x_0(m) \) the \( m \) individuals in the population \( x_0 \). We denote by \( f_0^* \) the maximal fitness of the individuals in \( x_0 \) and by \( \bar{f}_0 \) their mean fitness, i.e.,

\[
f_0^* = \max_{1 \leq i \leq m} f(x_0(i)), \quad \bar{f}_0 = \frac{1}{m} \sum_{1 \leq i \leq m} f(x_0(i)).
\]
The disordered regime. We consider the fitness function $f$ defined by

$$\forall u \in \{0, 1\}^\ell \quad f(u) = \begin{cases} 2 & \text{if } u = 1 \cdots 1 \\ 1 & \text{otherwise} \end{cases}$$

This corresponds to the sharp peak landscape. The chromosome $1 \cdots 1$ is called the Master sequence. We start the genetic algorithm from the population $x_0$ containing one Master sequence $1 \cdots 1$ and $m - 1$ copies of the chromosome $0 \cdots 0$.

**Theorem 3.1** Let $\pi < 1$ be fixed. We suppose that the parameters are set so that $\ell = m$ and $(f_0^*/f_0)(1 - p_C)(1 - p_M)^\ell = \pi$. There exist strictly positive constants $\kappa, \beta, m_0$, which depend on $\pi$ only, such that, for the genetic algorithm starting from $x_0$, for any $m \geq m_0$,

$$P \left( \text{before generation } \kappa \ln m, \text{ the Master sequence disappears} \right) \geq 1 - \frac{1}{\sqrt{m}}.$$

The quasispecies regime. We consider an arbitrary fitness function $f$ and we start the genetic algorithm from an arbitrary population $x_0$.

**Theorem 3.2** Let $\pi > 1$ be fixed. We suppose that the parameters are set so that $(f_0^*/f_0)(1 - p_C)(1 - p_M)^\ell = \pi$. There exist strictly positive constants $\kappa, p^*$, which depend on $\pi$ and the ratio $f_0^*/f_0$ only, such that, for the genetic algorithm starting from $x_0$, for any $\ell, m \geq 1$,

$$P \left( \text{before generation } \kappa \ln m, \text{ the mean fitness becomes } \geq \sqrt{\pi f_0^*} \right) \geq p^*.$$

4 The disordered regime

The proof of theorem 3.1 follows closely the proof of theorem 2.1 in [1]. Thus we do not reproduce the full argument here. The proof has two main steps. First we define a process $(T_n)_{n \in \mathbb{N}}$ which counts the number of descendants of the Master sequence in generation $n$. We show that, as long as $T_n \leq m^{1/4}$, the process $(T_n)_{n \in \mathbb{N}}$ is stochastically dominated by a supercritical Galton–Watson process. Next we define a process $(N_n^*)_{n \in \mathbb{N}}$ which counts the number of Master sequences present in generation $n$. Note that $N_n^*$ is in general smaller than $T_n$, because of the mutations and the crossovers. Indeed a chromosome might have an ancestor which is a Master sequence and be very different from it. We show then that, as long as $T_n \leq m^{1/4}$, the process $(N_n^*)_{n \in \mathbb{N}}$ is stochastically dominated by a subcritical Galton–Watson process. We finally use a classical argument from
the theory of branching processes to prove that this subcritical Galton–Watson process becomes extinct before generation $\kappa \ln m$ with probability larger than $1 - 1/m^\beta$. The computations are tedious, because we need to control the probabilities of obtaining a Master sequence when applying the various genetic operators, and the crossover creates correlations between pairs of adjacent chromosomes. The only point we have to adapt with respect to the case of ranking selection concerns the value of the probability of selecting the Master sequence or an individual which is a descendant of the Master sequence. More precisely, the probability that an individual in generation $n$ is a descendant of the Master sequence, given that there are $T_{n-1}$ such descendants in generation $n-1$, is bounded above by

$$1 - \left( \frac{m - T_{n-1}}{2T_{n-1} + m - T_{n-1}} \right)^2 = \frac{4mT_{n-1}}{(m + T_{n-1})^2}.$$ 

This bound is obtained by considering the case where the $T_{n-1}$ descendants of the Master sequence are themselves Master sequences, and by computing the probability that at least one of the two parents is a Master sequence. Defining

$$\tau_1 = \inf \{ n \geq 1 : T_n > m^{1/4} \},$$

and using the same notation than in [1], we have then

$$P(M_n(2) = 1, \tau_1 \geq n \mid X_{n-1}, M_{n-1}, \ldots, X_0, M_0) \leq 1_{\{\tau_1 \geq n\}} \frac{4mT_{n-1}}{(m + T_{n-1})^2} \leq 1_{\{\tau_1 \geq n\}} \frac{8T_{n-1}}{m}.$$ 

This is the key estimate to prove that $(T_n)_{n \in \mathbb{N}}$ is stochastically dominated by a supercritical Galton–Watson process. Later in the proof, we bound the probability that an individual in generation $n$ has a parent which is a Master sequence, given that there are $N^*_n$ Master sequences in generation $n-1$: for $\varepsilon > 0$ and for $m$ large enough,

$$1_{\{\tau \geq n\}} \left( F_m(m) - F_m(m - N^*_n) \right) \leq 1_{\{\tau \geq n\}} \frac{4mN^*_n}{(m + m^{1/4})^2} \leq 1_{\{\tau \geq n\}} \frac{4(1 + \varepsilon)}{m} N^*_n.$$

Here $\tau$ is an adequate stopping time introduced in the course of the argument. This is the key estimate to prove that $(N^*_n)_{n \in \mathbb{N}}$ is stochastically dominated by a subcritical Galton–Watson process. Besides these two estimates, the rest of the proof is essentially unchanged.
5 The quasispecies regime

We start the genetic algorithm with an initial population whose maximal fitness is equal to $f_0^*$ and whose mean fitness is equal to $\overline{f}_0$. We first define an auxiliary Markov chain $(N_n)_{n \geq 0}$ on the integers, whose transition matrix depends on the ratio $f_0^*/\overline{f}_0$. This Markov chain starts from $N_0 = 1$. We describe next its transition mechanism. Let $B_n$ be a random variable with law the binomial law $\mathcal{B}(m/2, 1 - p_C)$. Conditionally on $N_n = i$, the law of $N_{n+1}$ is the same as the law of the random variable

$$\sum_{k=1}^{2B_n} Y_k^i,$$

where the variables $Y_k^i$, $k \in \mathbb{N}$, $i \in \{0, \ldots, m\}$, are Bernoulli i.i.d. random variables (independent of $B_n$ as well) with parameter

$$\varepsilon_m(i) = \frac{if_0^*}{m\sqrt{\pi} \overline{f}_0} (1 - p_M)^{i/2}.$$

In particular, the null state is an absorbing state. To sum up, we have for $i \in \{0, \ldots, m\}$ and $j \in \{0, \ldots, m\}$,

$$P(N_{n+1} = j \mid N_n = i) = \sum_{b=0}^{m/2} \binom{m/2}{b} (1 - p_C)^b p_C^{m/2-b} \binom{2b}{j} \varepsilon_m(i)^j (1 - \varepsilon_m(i))^{2b-j}.$$

For $x$ a population, we define $N(x, f_0^*)$ as the number of chromosomes in $x$ whose fitness is larger than or equal to $f_0^*$:

$$N(x, f_0^*) = \text{card} \{ i \in \{1, \ldots, m\} : f(x(i)) \geq f_0^* \}.$$

We denote by $X_n$ the population at generation $n$ and by $X_n(1), \ldots, X_n(m)$ the $m$ individuals of $X_n$. We define a stopping time $\tau$ by

$$\tau = \inf \left\{ n \geq 1 : \frac{1}{m} \left( f(X_n(1)) + \cdots + f(X_n(m)) \right) \geq \sqrt{\pi} \overline{f}_0 \right\}.$$

Using a construction similar to the coupling built in [1], we can couple the process $(N(X_n, f_0^*))_{n \in \mathbb{N}}$ with the Markov chain $(N_n)_{n \in \mathbb{N}}$ in such a way that

$$\forall n \geq 0 \quad N(X_n, f_0^*) \geq N_n 1_{\tau \geq n}.$$

If $N(X_n, f_0^*) \geq m$, then necessarily

$$\frac{1}{m} \left( f(X_n(1)) + \cdots + f(X_n(m)) \right) \geq f_0^* \geq \sqrt{\pi} \overline{f}_0 > \sqrt{\pi} \overline{f}_0.$$
and thus $\tau < n$. The above coupling inequality implies therefore that

$$P(\tau < n) \geq P(\exists k \leq n \ N_k \geq m).$$

We study then the dynamics of the Markov chain $(N_n)_{n \geq 0}$. Our goal is to prove that, with positive probability (independent of $m$), this Markov chain will reach a value larger than $m$ before time $\kappa \ln m$. In proposition 5.3, we shall bound from below the probability of hitting a value larger than $m$ before time $\kappa \ln m$ and this will conclude the proof of theorem 3.2. In the next proposition, we give the key estimate, which shows that the Markov chain $(N_n)_{n \geq 0}$ is likely to grow geometrically until a value larger than $m$. The proof is elementary, in the sense that it relies essentially on two classical exponential inequalities (which are recalled in the appendix). This proof is an adaptation of the proof of proposition 6.7 in [1].

**Proposition 5.1** Let $\pi > 1$ be fixed. There exist $\rho > 1$, $c_0 > 0$, $m_0 \geq 1$, which depend on $\pi$ and the ratio $f_0^* / f_0$ only, such that: for any set of parameters $\ell, p_C, p_M$ satisfying $\pi = (f_0^* / f_0)(1 - p_C)(1 - p_M)^\ell$, we have

$$\forall m \geq m_0 \ \forall i \in \{1, \ldots, m\} \ \ P(N_{n+1} \leq \rho i \mid N_n = i) \leq \exp(-c_0 i).$$

**Proof.** Let $\varepsilon > 0$ be such that $\sqrt{\pi(1 - 2\varepsilon)} > 1$ and let

$$l(m, \varepsilon) = \left\lfloor \frac{m}{2} (1 - p_C)(1 - \varepsilon) \right\rfloor + 1 + \frac{m}{4}(1 - p_C)\varepsilon.$$

For $m$ large enough, we have

$$l(m, \varepsilon) < \frac{m}{2} (1 - p_C)(1 - \varepsilon) + 1 < \frac{m}{2}(1 - p_C).$$

Let $\rho$ be such that $1 < \rho < \sqrt{\pi(1 - 2\varepsilon)}$. We write

$$P(N_{n+1} \leq \rho i \mid N_n = i) = P\left(\sum_{k=1}^{2B_n} Y_k^i < \rho i\right) \leq P(B_n \leq l(m, \varepsilon)) + P\left(\sum_{k=1}^{2l(m, \varepsilon)} Y_k^i < \rho i\right).$$

We control the first probability with the help of Hoeffding’s inequality (see the appendix). The expected value of $B_n$ is $m(1 - p_C)/2 > l(m, \varepsilon)$, thus

$$P(B_n \leq l(m, \varepsilon)) \leq \exp\left(-\frac{2}{m} \left(\frac{m}{2} (1 - p_C) - l(m, \varepsilon)\right)^2\right).$$

Recall that $1 - p_C > f_0^* / f_0^*$. For $m$ large enough, we have

$$\frac{m}{2} (1 - p_C) - l(m, \varepsilon) \geq \frac{m}{2} (1 - p_C) \varepsilon - 1 \geq \frac{m\varepsilon f_0}{4 f_0^*} - 1 \geq \frac{m\varepsilon f_0}{8 f_0^*}.$$
It follows that, for $m$ large enough,

$$P(B_n \leq l(m, \varepsilon)) \leq \exp \left( -\frac{m}{32} \left( \frac{\varepsilon f_0}{f_0^*} \right)^2 \right).$$

Let us try to apply also Hoeffding’s inequality to the second probability. We get

$$P \left( \sum_{k=1}^{2l(m, \varepsilon)} Y_k^i < \rho i \right) \leq \exp \left( -\frac{1}{l(m, \varepsilon)} \left( 2l(m, \varepsilon) \varepsilon_m(i) - \rho i \right)^2 \right).$$

Now

$$2l(m, \varepsilon) \varepsilon_m(i) \geq \frac{2}{\pi} \left( 1 - p_C \right) \left( 1 - \varepsilon \right) \frac{i f_0^*}{m \sqrt{\pi} f_0} \left( 1 - p_M \right)^i = \left( 1 - \varepsilon \right) i \sqrt{\pi},$$

whence, using the hypothesis on $\rho$,

$$P \left( \sum_{k=1}^{2l(m, \varepsilon)} Y_k^i < \rho i \right) \leq \exp \left( -\frac{\pi \varepsilon^2 i^2}{m} \right).$$

This inequality becomes useful only when $i$ of order $\delta m$ for some $\delta > 0$. For smaller values of $i$, we must proceed differently in order to control this probability. Thus we decompose the sum into $i$ blocks and we use the Tchebytcheff exponential inequality. Each block follows a binomial law, and we bound the Cramér transform of each block by the Cramér transform of a Poisson law having the same mean. More precisely, we choose for the block size

$$b = \left\lfloor \frac{2l(m, \varepsilon) - \frac{m}{4} (1 - p_C) \varepsilon}{i} + 1 \right\rfloor,$$

and we define the sum associated to each block of size $b$:

$$\forall j \in \{1, \ldots, i\} \quad Y_j' = \sum_{k=b(j-1)+1}^{b j} Y_k^i.$$  

Notice that $Y_j'$ follows the binomial law with parameters $b, \varepsilon_m(i)$. We will next estimate from below the product $b \varepsilon_m(i)$. By the choice of $b$ and $l$, we have

$$b \geq \frac{1}{i} \left( 2l(m, \varepsilon) - \frac{m}{4} (1 - p_C) \varepsilon \right), \quad l(m, \varepsilon) \geq \frac{m}{2} \left( 1 - p_C \right) \left( 1 - \varepsilon \right),$$

whence

$$b \geq \frac{m}{i} (1 - p_C) (1 - \varepsilon)$$
and
\[ E(Y'_1) = b\varepsilon_m(i) \geq \sqrt{\pi}(1 - \varepsilon) > \rho. \]

Let \( \delta_0 > 0 \) be such that \( \delta_0 < (1 - p_C)\varepsilon/4 \). Suppose that \( i \leq \delta_0 m \). We have also that
\[
bi \leq 2l(m, \varepsilon) - \frac{m}{4}(1 - p_C)\varepsilon + \delta_0 m \leq 2l(m, \varepsilon).
\]

Using Tchebytcheff exponential inequality (see the appendix), we have then
\[
P\left( \sum_{k=1}^{2l(m,\varepsilon)} Y'_k \leq \rho i \right) \leq P\left( \sum_{k=1}^{bi} Y'_k \leq \rho i \right)
\leq P\left( \sum_{j=1}^{i} -Y'_j \geq -\rho i \right) \leq \exp \left( -i\Lambda^*_{-Y'_1}(-\rho) \right),
\]
where \( \Lambda^*_{-Y'_1} \) is the Cramér transform of \( -Y'_1 \). Let \( Y''_1 \) be a random variable following the Poisson law of parameter \( b\varepsilon_m(i) \). We shall use the following lemma to compare the Cramér transforms of \( -Y'_1 \) and \( -Y''_1 \).

**Lemma 5.2** Let \( p \in [0, 1] \) and let \( n \geq 1 \). Let \( X \) be a random variable following the binomial law \( B(n, p) \). Let \( Y \) be a random variable following the Poisson law \( P(np) \). For any \( \alpha \in \mathbb{R} \), we have \( \Lambda^*_{\alpha X} \geq \Lambda^*_{\alpha Y} \).

**Proof.** For any \( t \in \mathbb{R} \), we have
\[
\Lambda_X(t) = \ln E\left( \exp(tX) \right) = n \ln \left( 1 - p + p \exp(t) \right) \leq np(\exp(t) - 1).
\]

For any \( \alpha, t \in \mathbb{R} \),
\[
\Lambda_{\alpha X}(t) = \Lambda_X(\alpha t) \leq np(\exp(\alpha t) - 1).
\]

Thus, taking \( \lambda = np \), we conclude that
\[
\forall t \in \mathbb{R} \quad \Lambda_{\alpha X}(t) \leq \Lambda_{\alpha Y}(t).
\]

Taking the Fenchel–Legendre transform, we obtain that \( \Lambda^*_{\alpha X} \geq \Lambda^*_{\alpha Y} \). \( \square \)

By lemma 5.2, we have
\[
\Lambda^*_{-Y'_1}(-\rho) \geq \Lambda^*_{-Y''_1}(-\rho) = \rho \ln \left( \frac{\rho}{b\varepsilon_m(i)} \right) - \rho + b\varepsilon_m(i).
\]

The map
\[
\lambda \mapsto \rho \ln \left( \frac{\rho}{\lambda} \right) - \rho + \lambda
\]

10
is non-decreasing on $[\rho, +\infty[$ and $b_{\epsilon_m}(i) \geq \sqrt{\pi}(1 - \epsilon)$, thus
\[ \Lambda^* Y_i \geq \rho \ln \left( \frac{\rho}{\sqrt{\pi(1 - \epsilon)}} \right) - \rho + \sqrt{\pi(1 - \epsilon)}. \]

Let us denote by $c_0$ the righthand quantity. Then $c_0$ is positive and it depends only on $\rho, \pi, f_0^* / f_0$ and $\epsilon$. Finally, we have for $m$ large enough, $i \in \{ 1, \ldots, \lfloor \delta_0 m \rfloor \}$,
\[ \mathbb{P} \left( \sum_{k=1}^{2l(m, \epsilon)} Y_k^i \leq \rho i \right) \leq \exp(-c_0 i) \]
whence
\[ \mathbb{P} \left( N_{n+1} \leq \rho i \mid N_n = i \right) \leq \exp \left( -\frac{m}{32} \left( \frac{\epsilon_{0}}{f_0^*} \right)^2 \right) + \exp(-c_0 i). \]

For $i$ such that $\delta_0 m \leq i \leq m$, we had obtained
\[ \mathbb{P} \left( N_{n+1} \leq \rho i \mid N_n = i \right) \leq \exp \left( -\frac{m}{32} \left( \frac{\epsilon_{0}}{f_0^*} \right)^2 \right) + \exp \left( -\pi \epsilon^2 \delta_0^2 m \right). \]

Let $\eta \in ]0, 1[$ be small enough so that $\eta c_0 \leq \pi \epsilon^2 \delta_0^2$ and, for $m$ large enough,
\[ \exp \left( -\frac{m}{32} \left( \frac{\epsilon_{0}}{f_0^*} \right)^2 \right) \leq \exp \left( -\eta \frac{mc_0}{2} \right) \left( 1 - \exp \left( -\frac{c_0}{2} \right) \right). \]

For $m$ large enough and $i \in \{ 1, \ldots, \lfloor \delta_0 m \rfloor \}$, we have
\[ \mathbb{P} \left( N_{n+1} \leq \rho i \mid N_n = i \right) \leq \exp \left( -\eta \frac{ic_0}{2} \right) \left( 1 - \exp \left( -\frac{c_0}{2} \right) \right) + \exp \left( -\eta ic_0 \right) \leq \exp \left( -\eta \frac{ic_0}{2} \right) \leq \exp \left( -\eta \frac{ic_0}{2} \right). \]

For $m$ large enough and $\delta_0 m \leq i \leq m$, we have
\[ \mathbb{P} \left( N_{n+1} \leq \rho i \mid N_n = i \right) \leq \exp \left( -\eta \frac{mc_0}{2} \right) \left( 1 - \exp \left( -\frac{c_0}{2} \right) \right) + \exp \left( -\eta mc_0 \right) \leq \exp \left( -\eta \frac{ic_0}{2} \right). \]

These inequalities yield the claim of the proposition. \square

We define $\tau_m = \inf \{ n \geq 0 : N_n \geq m \}$. 

11
Proposition 5.3 Let $\pi > 1$ be fixed. There exist $\kappa > 0$ and $p^* > 0$ which depend on $\pi$ and the ratio $f_0^*/f_0$ only such that

$$\forall m \geq 1 \quad P(\tau_m \leq \kappa \ln m \mid N_0 = 1) \geq p^*.$$ 

Proof. Let us define

$$\tau_0 = \inf \{ n \geq 1 : N_n = 0 \}.$$ 

Recall that 0 is an absorbing state. Thus, if the hitting time of $m$ is finite, then necessarily, it is smaller than the hitting time of 0. It follows that

$$P(\tau_m \leq \kappa \ln m \mid N_0 = 1) = P(\tau_m \leq \kappa \ln m, \tau_m < \tau_0 \mid N_0 = 1).$$

It is annoying to work with a Markov chain which has an absorbing state, so we first get rid of this problem. We consider the modified Markov chain $(\tilde{N}_n)_{n \geq 0}$ which has the same transition probabilities as $(N_n)_{n \geq 0}$, except that we set the transition probability from 0 to 1 to be 1. The event we wish to estimate has the same probability for both processes, because they have the same dynamics outside of 0. So, from now onwards, we work with the Markov chain $(\tilde{N}_n)_{n \geq 0}$, which is irreducible. Let $\rho > 1$, $c_0 > 0$, $m_0 \geq 1$ be as given in proposition 5.1. For $k \geq 0$, let $T_k$ be the first time the process $(\tilde{N}_n)_{n \geq 0}$ hits $k$:

$$T_k = \inf \{ n \geq 0 : \tilde{N}_n = k \}.$$ 

Let $E$ be the event:

$$E = \{ \forall k \in \{ 1, \ldots, m \} \quad \tilde{N}_{T_k+1} \geq \rho k \}.$$ 

We claim that, on the event $E$, we have

$$\forall n \leq \tau_m \quad \tilde{N}_{n+1} \geq \rho \tilde{N}_n.$$ 

Let us prove this inequality by induction on $n$. We have $T_1 = 0$ and $\tilde{N}_1 > \rho \tilde{N}_0$, so that the inequality is true for $n = 0$. Suppose that the inequality has been proved until rank $n < \tau_m$, so that

$$\forall k \leq n \quad \tilde{N}_{k+1} \geq \rho \tilde{N}_k.$$ 

This implies in particular that

$$\tilde{N}_0 < \tilde{N}_1 < \ldots < \tilde{N}_n < m.$$ 

Suppose that $\tilde{N}_n = i$. The above inequality implies that $T_i = n$ and

$$\tilde{N}_{T_i+1} = \tilde{N}_{n+1} \geq \rho \tilde{N}_n.$$ 

12
so that the inequality still holds at rank \( n + 1 \). Iterating the inequality until time \( \tau_m - 1 \), we see that

\[
\tilde{N}_{\tau_m - 1} \geq \rho^{\tau_m - 1}.
\]

Moreover \( \tilde{N}_{\tau_m - 1} \leq m \), thus

\[
\tau_m \leq 1 + \frac{\ln m}{\ln \rho}.
\]

Let \( m_1 \geq 1 \) and \( \kappa > 0 \) be such that

\[
\forall m \geq m_1 \quad 1 + \frac{\ln m}{\ln \rho} \leq \kappa \ln m.
\]

The constants \( m_1, \kappa \) depend only on \( \rho \), and we have

\[
P(\tau_m \leq \kappa \ln m, \tau_m < \tau_0 \mid \tilde{N}_0 = 1) \geq P(\mathcal{E}).
\]

We shall use the following lemma to bound \( P(\mathcal{E}) \) from below. To avoid too small indices, we write \( T(i) \) instead of \( T_i \).

**Lemma 5.4** Let \( k \in \{1, \ldots, m\} \) and let \( i_1, \ldots, i_k \) be \( k \) distinct points of \( \{1, \ldots, m\} \). The random variables \( \tilde{N}_{T_{i_1} + 1}, \ldots, \tilde{N}_{T_{i_k} + 1} \) are independent.

**Proof.** We do the proof by induction over \( k \). For \( k = 1 \), there is nothing to prove. Let \( k \geq 2 \) and suppose that the result has been proved until rank \( k - 1 \). Let \( i_1, \ldots, i_k \) be \( k \) distinct points of \( \{1, \ldots, m\} \). Let \( j_1, \ldots, j_k \) be \( k \) points of \( \{1, \ldots, m\} \). Let us set

\[
T = \min \{ T(i_l) : 1 \leq l \leq k \}.
\]

We denote by \( (p(i, j))_{0 \leq i,j \leq m} \) the transition matrix of the Markov chain \((\tilde{N}_n)_{n \geq 0}\). Using the Markov property, we have

\[
P(\tilde{N}_{T(i_1)} + 1 = j_1, \ldots, \tilde{N}_{T(i_k)} + 1 = j_k)
= \sum_{1 \leq l \leq k} P(\tilde{N}_{T(i_l)} + 1 = j_l, T = T(i_l))
= \sum_{1 \leq l \leq k} P(\tilde{N}_{T(i_l)} + 1 = j_l, \tilde{N}_0 = j_l) P(T = T(i_l))
= \sum_{1 \leq l \leq k} p(i_l, j_l) P(\forall h \neq l \quad \tilde{N}_{T(i_h)} + 1 = j_h, \tilde{N}_0 = i_l) P(T = T(i_l)).
\]
We use the induction hypothesis:

\[ P(\forall h \neq l \mid \tilde{N}_{T(i_l)+1} = j_h | \tilde{N}_0 = j_l) = \prod_{h \neq l} p(i_h, j_h). \]

Reporting in the sum, we get

\[ P(\tilde{N}_{T(i_1)+1} = j_1, \ldots, \tilde{N}_{T(i_k)+1} = j_k) = \sum \prod_{1 \leq h \leq k} p(i_h, j_h) P(T = T(i_l)). = \prod_{1 \leq h \leq k} p(i_h, j_h). \]

This completes the induction step and the proof. □

Using lemma 5.4 and proposition 5.1, we obtain, for \( m \) larger than \( m_0 \) and \( m_1 \),

\[ P(\mathcal{E}) \geq \prod_{1 \leq k \leq m} P(\tilde{N}_{T_k+1} \geq \rho k) \]

\[ = \prod_{1 \leq k \leq m} \left( 1 - P(N_1 < \rho k | N_0 = k) \right) \]

\[ \geq \prod_{1 \leq k \leq m} \left( 1 - \exp(-c_0 k) \right) \geq \prod_{k=1}^{\infty} \left( 1 - \exp(-c_0 k) \right). \]

The last infinite product is converging. Let us denote its value by \( p_1 \). Let also

\[ p_2 = \min \left\{ P(\tau_m \leq \kappa \ln m | N_0 = 1) : m \leq \max(m_0, m_1) \right\}. \]

The value \( p_2 \) is positive and the inequality stated in the proposition holds with \( p^* = \min(p_1, p_2) \). □

**Appendix**

**Hoeffding’s inequality.** We state Hoeffding’s inequality for Bernoulli random variables [5]. Suppose that \( X \) is a random variable with law the binomial law \( B(n,p) \). We have

\[ \forall t < np \quad P(X < t) \leq \exp \left( -\frac{2}{n} (np - t)^2 \right). \]

**Tchebytcheff exponential inequality.** Let \( X_1, \ldots, X_n \) be i.i.d. random variables with common law \( \mu \). Let \( \Lambda \) be the Log–Laplace of \( \mu \), defined by

\[ \forall t \in \mathbb{R} \quad \Lambda(t) = \ln \left( \int_{\mathbb{R}} \exp(ts) \, d\mu(s) \right). \]
Let \( \Lambda^* \) be the Cramér transform of \( \mu \), defined by
\[
\forall x \in \mathbb{R} \quad \Lambda^*(x) = \sup_{t \in \mathbb{R}} \left( tx - \Lambda(t) \right).
\]
We suppose that \( \mu \) is integrable and we denote by \( m \) its mean, i.e., \( m = \int_{\mathbb{R}} x \, d\mu(x) \). We have then (see for instance [2])
\[
\forall x \geq m \quad P\left( \frac{1}{n}(X_1 + \cdots + X_n) \geq x \right) \leq \exp\left( -n \Lambda^*(x) \right).
\]
Let \( Y \) be a random variable following the Poisson law \( P(\lambda) \). For any \( t \in \mathbb{R} \), we have
\[
\Lambda_Y(t) = \ln E\left( \exp(tY) \right) = \ln \left( \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \exp(-\lambda + kt) \right) = \lambda \left( \exp(t) - 1 \right).
\]
For any \( \alpha, t \in \mathbb{R} \),
\[
\Lambda_{\alpha Y}(t) = \Lambda_Y(\alpha t) = \lambda \left( \exp(\alpha t) - 1 \right).
\]
Let us compute the Fenchel–Legendre transform \( \Lambda_{\alpha Y}^* \). By definition, for \( x \in \mathbb{R} \),
\[
\Lambda_{\alpha Y}^*(x) = \sup_{t \in \mathbb{R}} \left( tx - \lambda \left( \exp(\alpha t) - 1 \right) \right).
\]
The maximum is attained at \( t = (1/\alpha) \ln(x/(\lambda \alpha)) \), hence
\[
\Lambda_{\alpha Y}^*(x) = \frac{x}{\alpha} \ln \left( \frac{x}{\lambda \alpha} \right) - \frac{x}{\alpha} + \lambda.
\]

References


