Adv. Appl. Prob. **49,** 903–926 (2017)
*doi:*10.1017/apr.2017.26

THE QUASISPECIES REGIME FOR THE SIMPLE GENETIC ALGORITHM WITH ROULETTE WHEEL SELECTION

RAPHAËL CERF,* *École Normale Supérieure*

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 believe that the genetic algorithm operates best when this parameter is slightly larger than 1 and we prove two results supporting this belief. We consider the case of the simple genetic algorithm with the roulette wheel selection mechanism. We denote by ℓ the length of the chromosomes, m the population size, p_{C} the crossover probability, and p_{M} the mutation probability. Our results suggest that the mutation and crossover probabilities should be tuned so that, at each generation, the maximal fitness multiplied by $(1-p_{\mathrm{C}})(1-p_{\mathrm{M}})^{\ell}$ is greater than the mean fitness.

Keywords: Genetic algorithm; quasispecies; stochastic optimization

2010 Mathematics Subject Classification: Primary 60J20
Secondary 92D15

1. Introduction

We study the classical simple genetic algorithm with the roulette wheel selection mechanism, as described in [5] and [7]. For the reader unfamiliar with this algorithm, let us provide a brief description of this simple genetic algorithm. The reader familiar with it can jump directly to the next paragraph. 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 [5]. 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 generation $n+1$ from generation n:

Repeat:

- select two chromosomes from generation n;
- perform the crossover;
- perform the mutation;
- put the two resulting chromosomes in generation $n+1$;

Until there are m chromosomes in generation $n+1$.

When building generation $n+1$ from generation n, all the random choices are performed independently. We use the classical genetic operators, as in [5], which we recall briefly.

Received 12 May 2016; revision received 8 December 2016.
* Postal address: Département de Mathématiques et Applications, École Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France. Email address: raphael.cerf@gmail.com

Selection. We use roulette wheel selection with replacement. Let us suppose that the current population is $x = (x(1), \dots, x(m))$. The probability of selecting the ith chromosome $x(i)$ in population x is given by the selection distribution defined by

$$\mathbb{P}(\text{select } i\text{th chromosome in } x) = \frac{f(x(i))}{f(x(1)) + \cdots + f(x(m))}.$$

Crossover. We use the standard single point crossover and the crossover probability is denoted by p_{C}. More precisely, we draw a coin of parameter p_{C} to decide whether or not a crossover occurs. If a crossover occurs then we select randomly and uniformly a cutting site along the chromosome and we swap the final segments of the chromosomes, as in the following example:

$$\mathbb{P}\left(\begin{array}{c|c} 000\,011 & 011\,001 \\ 100\,110 & 001\,111 \end{array} \rightarrow \begin{array}{c|c} 000\,011 & 001\,111 \\ 100\,110 & 011\,001 \end{array} \right) = \frac{p_{\mathrm{C}}}{\ell - 1}.$$

Mutation. We use independent parallel mutation at each bit and the mutation probability is denoted by p_{M}. More precisely, each bit of the chromosome is transformed into the complementary bit with probability p_{M} and kept unchanged with probability $1 - p_{\mathrm{M}}$, and these decisions are independent. For instance, we have

$$\mathbb{P}(0000000 \rightarrow 0101000) = p_{\mathrm{M}}^2 (1 - p_{\mathrm{M}})^5.$$

Genetic algorithms have been used in practice for a myriad of problems, with varying successes. Moreover, many variants have been considered, which include different genetic mechanisms, as well as different coding methods. Yet a central problem to efficiently implement 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 vast 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 [9] for a recent review. Throughout this work, we focus on the simplest genetic algorithm but we anticipate that similar results might be proved for variants of the algorithm. For instance, our results are not restricted to binary strings and they hold for any finite alphabet. Similarly, we deal only with the one-point crossover, but our results depend essentially on the probability $1 - p_{\mathrm{M}}$ of not having a crossover, thus, they can be readily extended to other crossover mechanisms.

Our goal 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 π. We suggest that, at any generation, the various operators of the genetic algorithm should be controlled in order to ensure that π 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 π is much larger than 1, in order to avoid the premature convergence of the algorithm. The optimal situation is when the population retains the best-fit chromosomes and actively explores their neighborhoods. Ideally we would like to have a few copies of the best-fit chromosomes and a cloud of mutants descending from them. This is why we aim at tuning the parameters so that π is only slightly

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

larger than 1. An interesting attempt 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 chromosomes, and it might be that, even using elitism, the best-fit chromosomes are quickly forgotten during the search and none of them have a chance to become the most recent common ancestor.

An advantage with the parameter π 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 the case of large populations. This kind of analysis has been previously conducted for the simple genetic algorithm with ranking selection [2]. Here we attempt to extend this analysis to the simple genetic algorithm with roulette wheel selection. This task turns 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 chromosomes are typically lost, or in a quasispecies regime, where the best-fit chromosomes survive and invade a positive fraction of the population. Our results have their roots in the quasispecies theory developed by Eigen *et al.* [4]. We refer the reader to the introduction of [2] 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 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 show that, in the limit of large populations, the dynamics of the genetic algorithm depends in a critical way on the parameter

$$\pi = \left(\frac{f_0^*}{\bar{f}_0}\right)(1 - p_{\mathrm{C}})(1 - p_{\mathrm{M}})^{\ell}.$$

- If $\pi < 1$ then the genetic algorithm might operate in a disordered regime: there exist positive constants β and κ 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 chromosome 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 κ and 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 at each generation, the mutation and crossover probabilities should be tuned so that

$$\text{maximal fitness} \times (1 - p_{\mathrm{C}})(1 - p_{\mathrm{M}})^{\ell} > \text{mean fitness}.$$

It seems, therefore, judicious to choose 'large' values of p_{M} and p_{C} compatible with the condition $\pi > 1$. In the generic situation where f_0^* is significantly larger than $\bar{f}_0$, this means

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

that the mutation probability should be of the order $1/\ell$; more precisely, the condition $\pi > 1$ implies that

$$\ell p_{\mathrm{M}} + p_{\mathrm{C}} < \ln\left(\frac{f_0^*}{\bar{f}_0}\right).$$

2. The results

We denote by x_0 the initial population and by $x_0(1), \dots, x_0(m)$ the m chromosomes in population x_0. We denote by f_0^* the maximal fitness of the chromosomes in x_0 and by $\bar{f}_0$ their mean fitness, i.e.

$$f_0^* = \max_{1 \le i \le m} f(x_0(i)), \qquad \bar{f}_0 = \frac{1}{m} \sum_{1 \le i \le m} f(x_0(i)).$$

We present two results to illustrate the contrasting behavior of the genetic algorithm when $\pi < 1$ and $\pi > 1$. For $k \ge 1$, we denote by f_k^* (respectively, $\bar{f}_k$) the maximal fitness (respectively, the mean fitness) of the chromosomes in the kth population generated by the genetic algorithm.

The disordered regime. We consider the fitness function f defined by, for all $u \in \{0, 1\}^\ell$,

$$f(u) = \begin{cases} 2 & \text{if } u = 1 \cdots 1, \\ 1 & \text{otherwise.} \end{cases}$$

This corresponds to the sharp peak landscape. The chromosome $1 \cdots 1$ is called the master sequence. We start the genetic algorithm from population x_0 containing one master sequence $1 \cdots 1$ and $m - 1$ copies of the chromosome $0 \cdots 0$. Thus, the optimal chromosome is already present in the population. Our goal is to study its influence on the evolution of the population. This is a crude model for the following scenario: the genetic algorithm has been stuck for a long time, and suddenly, by chance, a chromosome with a superior fitness is found; is this new chromosome likely to influence the whole population or will it disappear? The next theorem describes a situation where the mean fitness of the population is unlikely to increase despite the presence of a very well fit chromosome.

Theorem 2.1. *Let $\pi < 1$ be fixed. We suppose that the parameters are set so that $\ell = m$ and $(f_0^*/\bar{f}_0)(1 - p_{\mathrm{C}})(1 - p_{\mathrm{M}})^\ell = \pi$. There exist strictly positive constants κ, β, and m_0, which depend on π only, such that, for the genetic algorithm starting from x_0, for any $m \ge m_0$,*

$$\mathbb{P}\left(\text{there exist } k \le \kappa \ln m,\ f_k^* = 1, \text{ for all } k \le \kappa \ln m,\ \bar{f}_k < \bar{f}_0\left(1 + \frac{1}{\sqrt{m}}\right)\right) \ge 1 - \frac{1}{m^\beta}.$$

The quasispecies regime. We consider an arbitrary nonnegative fitness function f and we start the genetic algorithm from a population x_0 such that $f_0^* > \bar{f}_0$. The next theorem describes a situation where the mean fitness of the population is likely to increase thanks to the influence of the best-fit chromosome.

Theorem 2.2. *Let $\pi > 1$ be fixed. We suppose that the parameters are set so that $(f_0^*/\bar{f}_0)(1 - p_{\mathrm{C}})(1 - p_{\mathrm{M}})^\ell = \pi$. There exist strictly positive constants κ and p^*, which depend on π and the ratio $f_0^*/\bar{f}_0$ only, such that, for the genetic algorithm starting from x_0, for any $\ell, m \ge 1$,*

$$\mathbb{P}(\text{for all } k \le \kappa \ln m,\ f_k^* \ge f_0^*, \text{there exist } \kappa \le \kappa \ln m,\ \bar{f}_k \ge \sqrt{\pi}\ \bar{f}_0) \ge p^*.$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

3. The disordered regime

In this section we will prove Theorem 2.1. 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 $T_n \le m^{1/4}$, 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 $T_n \le m^{1/4}$, the process $(N_n^*)_{n\in\mathbb{N}}$ is stochastically dominated by a subcritical Galton–Watson process. The bound on $(N_n^*)_{n\in\mathbb{N}}$ relies on the previous bound on $(T_n)_{n\in\mathbb{N}}$. We finally invoke 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.

Let us start with the precise proof. We start the genetic algorithm from population x_0 containing one master sequence $1\cdots 1$ and $m-1$ copies of the chromosome $0\cdots 0$. Let $\pi < 1$ be fixed. Throughout the proof, we suppose that ℓ, p_{C}, and p_{M} satisfy $\ell = m$ and

$$2(1-p_{\mathrm{C}})(1-p_{\mathrm{M}})^{\ell} = \pi.$$

We denote by X_n the population at generation n and by T_n the number of descendants of the initial master sequence present in X_n. To build generation $n+1$, we select (with replacement) m chromosomes from population X_n. Let us denote by A_n the number of chromosomes selected in X_n which are a descendant of the initial master sequence. Each of these chromosomes is the parent of two chromosomes in generation $n+1$ (because of the crossover operator). Thus, we can bound T_{n+1} from above by $2A_n$. Conditionally on T_n, the distribution of A_n is stochastically dominated by a binomial with parameters m and

$$\frac{2T_n}{2T_n + m - T_n} \le \frac{2T_n}{m}.$$

Thus, conditionally on T_n, the distribution of T_{n+1} is stochastically dominated by the binomial distribution $2\mathcal{B}(m, 2T_n/m)$, which we write as

$$T_{n+1} \preceq 2\mathcal{B}\left(m, \frac{2}{m}T_n\right).$$

The symbol '$\preceq$' means stochastic domination (see Appendix A). We define

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

and we will compute estimates which hold until time τ_1. So we fix $n \ge 1$ and we condition on the event that $\tau_1 > n$. There exists $t_0 > 0$ such that, for $0 < t < t_0$, we have $\ln(1-t) \ge -2t$. Therefore, for large enough m so that $2m^{-3/4} < t_0$, we have

$$\left(1 - \frac{2}{m}T_n\,\mathbf{1}_{\{\tau_1 > n\}}\right)^m \ge \exp(-4T_n\,\mathbf{1}_{\{\tau_1>n\}}).$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

We denote by $\mathcal{P}(\lambda)$ the Poisson law of parameter λ. By Lemma A.2, we conclude from this inequality that

$$2\mathcal{B}\left(m, \frac{2}{m} T_n \mathbf{1}_{\{\tau_1 > n\}}\right) \preceq 2\mathcal{P}(4T_n \mathbf{1}_{\{\tau_1 > n\}}).$$

Therefore,

$$T_{n+1} \mathbf{1}_{\{\tau_1 \geq n+1\}} \preceq \sum_{k=1}^{T_n \mathbf{1}_{\{\tau_1 > n\}}} V_k,$$

where the random variables $(V_k)_{k \geq 1}$ are independent and identically distributed (i.i.d.) with distribution twice the Poisson law of parameter 4. Let $(Z_n)_{n \in \mathbb{N}}$ be a Galton–Watson process starting from $Z_0 = 1$ with reproduction law $2\mathcal{P}(4)$. We conclude from the previous inequality that, for all $n \geq 0$,

$$T_n \mathbf{1}_{\{\tau_1 \geq n\}} \preceq Z_n.$$

We denote by $X_n(1), \ldots, X_n(m)$ the m chromosomes of population X_n. Let N_n^* be the number of master sequences present in the population at time n, i.e. for all $n \geq 0$,

$$N_n^* = \operatorname{card}\{i \in \{1, \ldots, m\} \colon X_n(i) = 1 \cdots 1\}.$$

We want to control N_{n+1}^* conditionally on the knowledge of N_n^* and T_n. A difficulty is that the crossover operator creates correlations between the chromosomes of X_{n+1}. However, conditionally on X_n, the pairs of consecutive chromosomes

$$(X_{n+1}(1), X_{n+1}(2)), \ldots, (X_{n+1}(m-1), X_{n+1}(m))$$

are i.i.d. Therefore, we can write N_{n+1}^* as the sum

$$N_{n+1}^* = \sum_{i=1}^{m/2} Y_i,$$

where Y_i is the number of master sequences in the ith pair $(X_{n+1}(2i-1), X_{n+1}(2i))$. Our strategy consists of estimating the conditional distribution of the Y_i, knowing population X_n. Conditionally on X_n, the random variables Y_i, $1 \leq i \leq m/2$, are i.i.d. with values in $\{0, 1, 2\}$, yet the computations are a bit lengthy and tedious because we have to consider all the possible cases, depending on whether the parents of $X_{n+1}(2i-1), X_{n+1}(2i)$ do or do not belong to the progeny of the initial master sequence. So let us focus on one pair of chromosomes, for instance, the first one $(X_{n+1}(1), X_{n+1}(2))$.

We have to estimate all the conditional probabilities

$$\mathbb{P}(\text{there are zero, one, or two master sequences in } (X_{n+1}(1), X_{n+1}(2)) \mid X_n).$$

To control these probabilities, we introduce the time τ_2, when a mutant, not belonging to the progeny of the initial master sequence, has at least $\sqrt{\ell}$ number of 1s. We set

$$\tau_2 = \inf\{n \geq 1 \colon \text{a chromosome of } X_n \text{ not in the progeny of the initial master sequence has } \sqrt{\ell} \text{ number of 1s}\}.$$

Let $\lambda > 0$ be such that $\pi/2 \geq \exp(-\lambda)$. We have

$$(1 - p_{\mathrm{M}})^{\ell} \geq \frac{\pi}{2(1 - p_{\mathrm{C}})} \geq \frac{\pi}{2} \geq \exp(-\lambda).$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Note that λ depends only on π and not on ℓ or p_{M}. By Lemma A.2, the binomial law $\mathcal{B}(\ell, p_{\mathrm{M}})$ is then stochastically dominated by the Poisson law $\mathcal{P}(\lambda)$. We will use repeatedly the bound on the tail of the Poisson law given in Lemma A.3, i.e. for all $t \geq \lambda$,

$$\mathbb{P}(\text{a given chromosome undergoes at least } t \text{ mutations from one generation to the next}) \leq \left(\frac{\lambda \mathrm{e}}{t}\right)^t.$$

When using this bound, the value of t will be a function of ℓ. We will always take ℓ large enough, so that the value of t will be larger than λ. We prove next a bound on τ_2.

Lemma 3.1. *For $m \geq 2$ and for large enough ℓ, we have*

$$\mathbb{P}\left(\tau_2 \leq \tfrac{1}{5} \ln \ell\right) \leq 1 - \exp(-m \exp(-\ell^{1/4})).$$

Proof. If $\tau_2 < n$ then, before time n, a chromosome has been created with at least $\sqrt{\ell}$ number of 1s, and whose genealogy does not contain the initial master sequence. We shall compute an upper bound on the number of 1s appearing in the genealogy of such a chromosome at generation n. Let us define D_n as the maximum number of 1s in a chromosome of generation n, which does not belong to the progeny of the initial master sequence. These 1s must have been created by mutation. Let us consider a chromosome of generation $n+1$, which does not belong to the progeny of the initial master sequence. The number of 1s in each of its two parents was at most D_n. After crossover between these two parents, the number of 1s was at most $2D_n$. After mutation, the number of 1s was at most

$$D_{n+1} \leq 2D_n + \max\{\text{number of mutations occurring on a chromosome between generation } n \text{ and } n+1\}.$$

We first control the last term. Let $n \geq 1$ and let us define the event $\mathcal{E}(n)$ by

$$\mathcal{E}(n) = \{\text{until generation } n, \text{ during the mutation process, the number of mutations occurring on a given chromosome is at most } \ell^{1/4}\}.$$

We have

$$\mathbb{P}(\mathcal{E}(n)) = (1 - \mathbb{P}(\text{a given chromosome undergoes more than } \ell^{1/4} \text{ mutations}))^{mn}.$$

Using the bound given in Lemma A.3, we obtain, for $\ell^{1/4} > \lambda$,

$$\mathbb{P}(\mathcal{E}(n)) \geq \left(1 - \left(\frac{\lambda \mathrm{e}}{\ell^{1/4}}\right)^{\ell^{1/4}}\right)^{mn},$$

whence, for large enough ℓ,

$$\mathbb{P}(\mathcal{E}(n)) \geq \exp(-mn \exp(-\ell^{1/4})).$$

Suppose that the event $\mathcal{E}(n)$ occurs. We have, for all $k \in \{0, \ldots, n-\}$,

$$D_{k+1} \leq 2D_k + \ell^{1/4}.$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Dividing by 2^{k+1} and summing from $k = 0$ to $n - 1$, we obtain

$$D_n \le 2^n \sum_{k=0}^{n-1} \frac{\ell^{1/4}}{2^{k+1}} \le 2^n \ell^{1/4}.$$

Therefore, if $2^n < \ell^{1/4}$ and if the event $\mathcal{E}(n)$ occurs then $\tau_2 > n$. Taking $n = (\ln \ell)/5$, we obtain the estimate stated in the lemma. □

We recall that

$$\tau_1 = \inf\{n \ge 1 \colon T_n > m^{1/4}\}.$$

We set also

$$\tau_0 = \inf\{n \ge 1 \colon N_n^* = 0\}.$$

We shall compute a bound on N_n^* until time $\tau = \min(\tau_0, \tau_1, \tau_2)$. Our goal is to show that, for large enough m, the process $(N_n^* \mathbf{1}_{\{\tau \ge n\}})_{n \in \mathbb{N}}$ is stochastically dominated by a subcritical Galton–Watson process. So let $n \ge 0$ and let us suppose that $\tau > n$ and that we know population X_n. We estimate the probability that exactly one master sequence is present in $(X_{n+1}(1), X_{n+1}(2))$. We envisage different scenarios, depending on the number of descendants of the initial master sequence among the two parents of these chromosomes.

First scenario. The two parents are descendants of the master sequence. The probability of selecting two such parents is bounded from above by

$$\left(\frac{2T_n}{2T_n + m - T_n}\right)^2 \le \left(\frac{2T_n}{m}\right)^2 \le \frac{4}{m\sqrt{m}}.$$

Second scenario. Exactly one of the parents is a descendant of the master sequence and a crossover has occurred. The total number of 1s present in the parents is at most $\ell + \sqrt{\ell}$. After crossover, the probability that one of the two resulting chromosomes has at least $\ell - \sqrt{\ell}$ number of 1s is less than $4/\sqrt{\ell}$. Indeed, this can happen only if, either on the left of the cutting site, or on its right, there are at most $\sqrt{\ell}$ number of 0s. The most favorable situation is when all the 1s are at the end or at the beginning of the chromosome that is not a descendant of the master sequence, in which case we have $2\sqrt{\ell}$ cutting sites which leads to the desired result. Otherwise, both chromosomes after crossover have at least $\sqrt{\ell}$ number of 0s, and the probability to transform these 0s into 1s through mutations is less than $(\lambda \mathrm{e}/\sqrt{\ell})^{\sqrt{\ell}}$. We conclude that the probability of this scenario is bounded from above by

$$\left(\frac{4}{\sqrt{\ell}} + 2\left(\frac{\lambda \mathrm{e}}{\sqrt{\ell}}\right)^{\sqrt{\ell}}\right) \frac{2N_n^*}{2N_n^* + m - N_n^*}.$$

Third scenario. Exactly one of the parents is a descendant of the master sequence and no crossover has occurred. A master sequence can be created from the chromosome not in the progeny of the initial master sequence, this would require $\ell - \sqrt{\ell}$ mutations, and the corresponding probability is bounded from above by

$$\left(\frac{\lambda \mathrm{e}}{\ell - \sqrt{\ell}}\right)^{\ell - \sqrt{\ell}}.$$

The other possibility is that a master sequence is obtained from the chromosome belonging to the progeny of the initial master sequence. This chromosome was either a master sequence,

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

in which case the replication has to be exact, or it was differing from the master sequence, in which case some mutations are required. The corresponding probability is bounded from above by

$$2(1-p_{\mathrm{C}})((1-p_{\mathrm{M}})^{\ell}+p_{\mathrm{M}})\frac{2N_n^*}{2N_n^*+m-N_n^*}.$$

Fourth scenario. None of the parents is a descendant of the master sequence. Until time τ_2, the chromosomes which are not descendants of the master sequence have at most $\sqrt{\ell}$ number of 1s. To create a master sequence starting from two such parents requires at least $\ell-2\sqrt{\ell}$ mutations. The corresponding probability is bounded from above by

$$2\left(\frac{\lambda \mathrm{e}}{\ell-2\sqrt{\ell}}\right)^{\ell-2\sqrt{\ell}}.$$

Combining the previous inequalities, we conclude that

$$\begin{aligned}&\mathbb{P}(\text{there is exactly one master sequence present in } (X_{n+1}(1), X_{n+1}(2)) \mid T_n,\ N_n^*)\\ &\quad\le \frac{4}{m\sqrt{m}}+\frac{4}{\sqrt{\ell}}+\left(\frac{4}{\sqrt{\ell}}+2\left(\frac{\lambda \mathrm{e}}{\sqrt{\ell}}\right)^{\sqrt{\ell}}\right)\frac{2N_n^*}{2N_n^*+m-N_n^*}\\ &\qquad+2(1-p_{\mathrm{C}})((1-p_{\mathrm{M}})^{\ell}+p_{\mathrm{M}})\frac{2N_n^*}{2N_n^*+m-N_n^*}+2\left(\frac{\lambda \mathrm{e}}{\ell-2\sqrt{\ell}}\right)^{\ell-2\sqrt{\ell}}.\end{aligned}$$

We rewrite the previous inequalities in the $\ell=m$ case and for large m. Since $2(1-p_{\mathrm{M}})^m \ge \pi$, then $p_{\mathrm{M}} \le -(1/m)\ln(\pi/2)$. Let $\varepsilon>0$ be such that $\pi(1+5\varepsilon)<1$. For large enough m and $n<\tau$, we have

$$\begin{aligned}&\mathbb{P}(\text{there is exactly one master sequence present in } (X_{n+1}(1), X_{n+1}(2)) \mid T_n, N_n^*)\\ &\quad\le \frac{2}{m}\pi(1+\varepsilon)N_n^*.\end{aligned}$$

Similar computations yield that there exists a positive constant c such that, for large enough m and $n<\tau$,

$$\mathbb{P}(\text{both } X_{n+1}(1), X_{n+1}(2) \text{ are master sequences} \mid T_n, N_n^*) \le \frac{c}{m^{3/2}}.$$

Returning to the initial equality for N_{n+1}^*, we conclude that, for large enough m, the law of $N_{n+1}^*\,\mathbf{1}_{\{\tau\ge n+1\}}$ is stochastically dominated by the sum of two independent binomial random variables as follows:

$$N_{n+1}^*\,\mathbf{1}_{\{\tau\ge n+1\}} \preceq \mathcal{B}\left(\frac{m}{2},\frac{2}{m}\pi(1+2\varepsilon)N_n^*\,\mathbf{1}_{\{\tau\ge n\}}\right)+2\mathcal{B}\left(\frac{m}{2},\frac{c}{m^{3/2}}\right).$$

For large m, these two binomial laws are, in turn, stochastically dominated by two Poisson laws. More precisely, for large enough m,

$$\left(1-\frac{2}{m}\pi(1+2\varepsilon)N_n^*\,\mathbf{1}_{\{\tau\ge n\}}\right)^{m/2} \ge \exp(-\pi(1+3\varepsilon)N_n^*\,\mathbf{1}_{\{\tau\ge n\}}),$$

$$(1-cm^{-3/2})^{m/2} \ge \exp(-\varepsilon).$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Lemma A.2 yields that

$$N^*_{n+1}\,\mathbf{1}_{\{\tau\geq n+1\}} \preceq \mathcal{P}(\pi(1+3\varepsilon)N^*_n\,\mathbf{1}_{\{\tau\geq n\}}) + 2\mathcal{P}(\varepsilon).$$

The point is that we have removed the variable m in the upper bound, so we are now in position to compare $N^*_n\,\mathbf{1}_{\{\tau\geq n\}}$ with a Galton–Watson process. Let $(Y'_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with law $\mathcal{P}(\pi(1+3\varepsilon))$, let $(Y''_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with law $\mathcal{P}(\varepsilon)$, both sequences being independent. The previous stochastic inequality can be written as

$$N^*_{n+1}\,\mathbf{1}_{\{\tau\geq n+1\}} \preceq \left(\sum_{k\geq 1}^{N^*_n\mathbf{1}_{\{\tau\geq n\}}} Y'_k\right) + 2Y''_1.$$

This implies further that

$$N^*_{n+1}\,\mathbf{1}_{\{\tau\geq n+1\}} \preceq \sum_{k\geq 1}^{N^*_n\mathbf{1}_{\{\tau\geq n\}}} (Y'_k + 2Y''_k). \tag{3.1}$$

Let ν^* be the law of $Y'_1 + 2Y''_1$ and let $(Z^*_n)_{n\geq 0}$ be a Galton–Watson process starting from $Z_0 = 1$ with reproduction law ν^*. We prove finally that, for large enough m, for all $n \geq 0$,

$$N^*_n\,\mathbf{1}_{\{\tau\geq n\}} \preceq Z^*_n.$$

We suppose that m is large enough so that the stochastic inequality (3.1) holds and we proceed by induction on n. For $n = 0$, we have

$$N^*_0\,\mathbf{1}_{\{\tau\geq 0\}} = 1 \leq Z^*_0 = 1.$$

Let $n \geq 0$ and suppose that the inequality holds at rank n. Inequality (3.1) yields

$$N^*_{n+1}\,\mathbf{1}_{\{\tau\geq n+1\}} \preceq \sum_{k\geq 1}^{N^*_n\mathbf{1}_{\{\tau\geq n\}}} (Y'_k + 2Y''_k) \preceq \sum_{k\geq 1}^{Z^*_n} (Y'_k + 2Y''_k) = Z^*_{n+1}.$$

Thus, the inequality holds at rank $n+1$ and the induction is completed. Moreover, we have

$$\mathbb{E}(\nu^*) = \mathbb{E}(Y'_1 + 2Y''_1) = \pi(1+5\varepsilon) < 1.$$

Thus, the Galton–Watson process $(Z^*_n)_{n\geq 0}$ is subcritical.

We now complete the proof of Theorem 2.1. Let $\kappa, c_1 > 0$ be constants associated to the Galton–Watson process $(Z_n)_{n\geq 0}$ as in Proposition A.1. We suppose that $\kappa < \frac{1}{5}$, so that we can use the estimate of Lemma 3.1. Let $c > 0$ be a constant associated to the subcritical Galton–Watson process $(Z^*_n)_{n\geq 0}$ as in Lemma A.5. We have

$$\begin{aligned}
\mathbb{P}(\tau_0 > \kappa \ln m) &\leq \mathbb{P}(\tau_0 > \kappa \ln m, \tau < \kappa \ln m) + \mathbb{P}(N^*_{\lfloor \kappa \ln m\rfloor} > 0, \tau \geq \kappa \ln m) \\
&\leq \mathbb{P}(\tau_1 < \kappa \ln m) + \mathbb{P}(\tau_2 < \kappa \ln m) + \mathbb{P}(Z^*_{\lfloor \kappa \ln m\rfloor} > 0) \\
&\leq \frac{1}{m^{c_1}} + 1 - \exp(-m\exp(-m^{1/4})) + \exp(-c^*\lfloor \kappa \ln m\rfloor).
\end{aligned}$$

This inequality yields the estimate stated in Theorem 2.1.

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

4. The quasispecies regime

In this section we will prove Theorem 2.2. We start the genetic algorithm with an initial population whose maximal fitness is equal to f_0^* and whose mean fitness is equal to $\bar{f}_0$. For a population $x = (x(1), \dots, x(m))$, we define $N(x, f_0^*)$ as the number of chromosomes in x whose fitness is larger than or equal to f_0^*, i.e.

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

We denote by X_n the population at generation n and by $X_n(1), \dots, X_n(m)$ the m chromosomes of X_n. We define a stopping time $\bar{\tau}$ by

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

Our goal is to control the time $\bar{\tau}$, more precisely we would like to prove that $\bar{\tau}$ is less than $\kappa \ln m$ with high probability. Unfortunately, the process $(N(X_n, f_0^*))_{n\geq 0}$ is very complicated, it is not even a Markov process. Our strategy is to construct an auxiliary Markov chain which is considerably simpler and which bounds $(N(X_n, f_0^*))_{n\geq 0}$ from below until time $\bar{\tau}$. The production of chromosomes with fitness larger than or equal to f_0^* from one generation to the next can be decomposed into two distinct mechanisms:

- chromosomes which are an exact copy of one of their parents;
- chromosomes which have undergone mutation or crossover events.

We will bound from below the process $(N(X_n, f_0^*))_{n\geq 0}$ by neglecting the second mechanism. The key point is that the law of the number of chromosomes created in generation $n+1$ through the first mechanism depends only on the value $N(X_n, f_0^*)$ and not on the detailed composition of the population at time n. Therefore, we are able to obtain a lower process which is a Markov chain. We denote this process by $(N_n)_{n\geq 0}$. We proceed next to its precise definition. Suppose that in generation n, we have i chromosomes of fitness larger than or equal to f_0^*, and that the mean fitness is still below $\sqrt{\pi}\, \bar{f}_0$, that is, we condition on the event $N(X_n, f_0^*) = i$ and $\bar{\tau} > n$. Let us look at the first pair of chromosomes of generation $n+1$. The probability to select from generation n a chromosome of fitness larger than or equal to f_0^* is at least $i f_0^*/(m\sqrt{\pi}\, \bar{f}_0)$. The probability that no crossover has occurred is $1 - p_{\text{C}}$. The probability that no mutation has occurred on a given chromosome is $(1 - p_{\text{M}})^\ell$. Thus, the probability that the first chromosome of generation $n+1$ is an exact copy of a chromosome of generation n having fitness larger than or equal to f_0^* is at least

$$\frac{i f_0^*}{m\sqrt{\pi}\, \bar{f}_0}(1 - p_{\text{C}})(1 - p_{\text{M}})^\ell.$$

However, the crossover creates correlations between adjacent chromosomes, so the distribution of N_{n+1} cannot be taken simply as a binomial law. Conditionally on the event that $N(X_n, f_0^*) = i$ and $\bar{\tau} > n$, a correct lower bound on $N(X_{n+1}, f_0^*)$ is given by the sum

$$\sum_{k=0}^{m/2} Z_k(Y_{2k-1} + Y_{2k}),$$

where $Z_1, \dots, Z_{m/2}$ are Bernoulli with parameter $1 - p_{\text{C}}$, and $Y_1, \dots, Y_m$ are Bernoulli with parameter

$$\varepsilon_m(i) = \frac{i f_0^*}{m\sqrt{\pi}\, \bar{f}_0}(1 - p_{\text{M}})^\ell$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

and they are all independent. The variable Z_k is 1 if there was no crossover between the chromosomes of the kth pair and 0 otherwise. The variable Y_k is 1 if the kth chromosome selected has fitness larger than or equal to f_0^* and it is not affected by any mutation. We obtain that, for $j \in \{0, \dots, m\}$,

$$\mathbb{P}(N(X_{n+1}, f_0^*) \geq j \mid N(X_n, f_0^*) = i, \bar{\tau} > n) \geq \mathbb{P}\left(\sum_{k=0}^{m/2} Z_k (Y_{2k-1} + Y_{2k}) \geq j\right).$$

We compute the right-hand side and we are led to define the transition matrix of the Markov chain $(N_n)_{n\geq 0}$ by setting, for $i, j \in \{0, \dots, m\}$,

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

The above inequality can then be written as, for $i, j \in \{0, \dots, m\}$,

$$\mathbb{P}(N(X_{n+1}, f_0^*) \geq j \mid N(X_n, f_0^*) = i, \bar{\tau} > n) \geq \mathbb{P}(N_{n+1} \geq j \mid N_n = i).$$

From Lemma A.1, this implies furthermore that, for any nondecreasing function $\phi\colon \mathbb{N} \to \mathbb{R}$, for $i \in \{0, \dots, m\}$,

$$\mathbb{E}(\phi(N(X_{n+1}, f_0^*)) \mid N(X_n, f_0^*) = i, \bar{\tau} > n) \geq \mathbb{E}(\phi(N_{n+1}) \mid N_n = i). \tag{4.1}$$

Let us focus a bit on the Markov chain $(N_n)_{n\geq 0}$. Its state space is $\{0, \dots, m\}$. The null state is an absorbing state because we neglect the mutations for producing chromosomes of fitness at least f_0^*. A key point to exploit inequality (4.1) is the following result.

Proposition 4.1. *The Markov chain $(N_n)_{n\geq 0}$ is monotone.*

Proof. The definition of monotone Markov chain is recalled in Appendix A (see Definition A.1). The easiest way to prove the monotonicity is to build an adequate coupling. For $n \in \mathbb{N}$ and $k \leq m/2$, let Z_k^n be a Bernoulli random variable with parameter $1 - p_{\mathrm{C}}$ and U_{2k-1}^n, U_{2k}^n be two random variables whose distribution is uniform over [0, 1]. We suppose that all the above random variables are independent. For $i \in \{0, \dots, m\}$, we define $N_0^i = i$ and, for all $n \geq 0$,

$$N_{n+1}^i = \sum_{k=0}^{m/2} Z_k^n (\mathbf{1}_{\{U_{2k-1}^n < \varepsilon_m(N_n)\}} + \mathbf{1}_{\{U_{2k}^n < \varepsilon_m(N_n)\}}).$$

This way all the chains $(N_n)_{n\geq 0}$, $i \in \{0, \dots, m\}$, are coupled and a straightforward induction yields that, for all $i \leq j$, $n \in \mathbb{N}$,

$$N_n^i \leq N_n^j.$$

This yields the desired conclusion. □

We are interested in the process $(N(X_n, f_0^*))_{n\geq 0}$ until time $\bar{\tau}$. In order to prove a convenient stochastic inequality, we will work with the process $(N_n^*)_{n\geq 0}$ defined by, for all $n \geq 0$,

$$N_n^* = \begin{cases} N(X_n, f_0^*) & \text{if } \bar{\tau} > n, \\ m & \text{if } \bar{\tau} \leq n. \end{cases}$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Proposition 4.2. *We suppose that the Markov chain $(N_n)_{n\in\mathbb{N}}$ starts from $N_0 = 1$. For any $n \geq 0$, we have the stochastic inequality*

$$(N_0^*, \ldots, N_n^*) \succeq (N_0, \ldots, N_n).$$

For the above statement, we work with the product order on $\mathbb{N}^{n+1}$, i.e.

$$(i_0, \ldots, i_n) \geq (j_0, \ldots, j_n) \iff i_0 \geq j_0, \ldots, i_n \geq j_n.$$

The stochastic domination inequality stated in Proposition 4.2 means that, for any nondecreasing function $\phi\colon \mathbb{N}^{n+1} \to \mathbb{R}^+$, we have

$$\mathbb{E}(\phi(N_0^*, \ldots, N_n^*)) \geq \mathbb{E}(\phi(N_0, \ldots, N_n)).$$

Proof of Proposition 2. We proceed by induction on n. For $n = 0$, we have

$$N_0^* = N(X_0, f_0^*) \geq 1 = N_0.$$

Suppose that the result has been proved until rank n for some $n \geq 0$. Let $\phi\colon \mathbb{N}^{n+2} \to \mathbb{R}^+$ be a nondecreasing function. We write

$$\begin{aligned}\mathbb{E}(\phi(N_0^*, \ldots, N_{n+1}^*)) = &\sum_{0\leq i_0,\ldots,i_n\leq m} \mathbb{P}(N_0^* = i_0, \ldots, N_n^* = i_n) \\ &\times \mathbb{E}(\phi(N_0^*, \ldots, N_{n+1}^*) \mid N_0^* = i_0, \ldots, N_n^* = i_n).\end{aligned}$$

Let $i_0, \ldots, i_n$ be fixed. Suppose first that $i_n < m$. The event $\{N_n^* = i_n\}$ implies that $\bar{\tau} > n$ and $N_n^* = N(X_n, f_0^*)$. The map

$$i \in \{0, \ldots, m\} \mapsto \phi(i_0, \ldots, i_n, i)$$

is nondecreasing. Using the stochastic inequality (4.1), we obtain

$$\begin{aligned}&\mathbb{E}(\phi(N_0^*, \ldots, N_{n+1}^*) \mid N_0^* = i_0, \ldots, N_n^* = i_n) \\ &\quad = \mathbb{E}(\phi(i_0, \ldots, i_n, N(X_{n+1}, f_0^*)) \mid N(X_0, f_0^*) = i_0, \ldots, N(X_n, f_0^*) = i_n) \\ &\quad \geq \mathbb{E}(\phi(i_0, \ldots, i_n, N_{n+1}) \mid N_n = i_n).\end{aligned}$$

Let us define a function $\psi\colon \mathbb{N}^{n+1} \to \mathbb{R}^+$ by setting

$$\psi(i_0, \ldots, i_n) = \mathbb{E}(\phi(i_0, \ldots, i_n, N_{n+1}) \mid N_n = i_n).$$

If $i_n = m$ then we also have

$$\mathbb{E}(\phi(N_0^*, \ldots, N_{n+1}^*) \mid N_0^* = i_0, \ldots, N_n^* = i_n) = \phi(i_0, \ldots, i_{n-1}, m, m) \geq \psi(i_0, \ldots, i_n).$$

From the previous inequalities, we conclude that

$$\begin{aligned}&\mathbb{E}(\phi(N_0^*, \ldots, N_{n+1}^*)) \\ &\quad \geq \sum_{0\leq i_0,\ldots,i_n\leq m} \mathbb{P}(N_0^* = i_0, \ldots, N_n^* = i_n)\psi(i_0, \ldots, i_n) = \mathbb{E}(\psi(N_0^*, \ldots N_n^*)).\end{aligned}$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Since the function ϕ is nondecreasing on $\mathbb{N}^{n+2}$ and since the Markov chain $(N_n)_{n\geq 0}$ is monotone (by Proposition 4.1), then the function ψ is also nondecreasing on $\mathbb{N}^{n+1}$. Now, the induction hypothesis yields that

$$\mathbb{E}(\psi(N_0^*, \dots, N_n^*)) \geq \mathbb{E}(\psi(N_0, \dots, N_n)) = \mathbb{E}(\phi(N_0, \dots, N_n, N_{n+1}))$$

and the induction step is completed.

If $N(X_n, f_0^*) > m/\sqrt{\pi}$ then necessarily

$$\frac{1}{m}(f(X_n(1)) + \cdots + f(X_n(m))) > \frac{1}{\sqrt{\pi}} f_0^* \geq \sqrt{\pi}\, \bar{f}_0$$

and, thus, $\bar{\tau} < n$. The above coupling inequality therefore implies that

$$\begin{aligned}\mathbb{P}(\bar{\tau} < n) &\geq \mathbb{P}\left(\text{there exists } k \leq n,\ N(X_k, f_0^*) > \frac{m}{\sqrt{\pi}}\right)\\ &\geq \mathbb{P}\left(\text{there exists } k \leq n,\ N_k > \frac{m}{\sqrt{\pi}}\right).\end{aligned}$$

We study next the dynamics of the Markov chain $(N_n)_{n\geq 0}$ on $\{0, \dots, m\}$. Our goal is to prove that, for some $\kappa > 0$, with a probability larger than a constant independent of m, this Markov chain will reach a value strictly larger than $m/\sqrt{\pi}$ before time $\kappa \ln m$. Let us explain briefly the heuristics for this result. The transition mechanism of the chain is built with the help of i.i.d. Bernoulli random variables, some of parameter $1 - p_{\mathrm{C}}$ and some of parameter $\epsilon_m(i)$, $i \in \{0, \dots, m\}$. The typical number of pairs of chromosomes with no crossover from one generation to another is $(1 - p_{\mathrm{C}})m/2$ and we can control accurately the deviations from this typical value. For i small compared to m, the parameter $\epsilon_m(i)$ is of the order of a constant multiplied by i/m, thus, conditionally on the event that $N_n = i$, the distribution of N_{n+1} is roughly the binomial law of parameters $m(1 - p_{\mathrm{C}})$ and a constant multiplied by i/m. In this regime, it can be approximated adequately by a Poisson law of parameter

$$m(1 - p_{\mathrm{C}})\epsilon_m(i) \sim i\sqrt{\pi}.$$

We conclude that, as long as N_n is small compared to m, we have

$$\mathbb{E}(N_{n+1}) \sim \sqrt{\pi}\, \mathbb{E}(N_n).$$

In the next proposition we derive a rigorous estimate, which indeed shows that the Markov chain $(N_n)_{n\geq 0}$ is likely to grow geometrically until a value larger than $m/\sqrt{\pi}$. The proof is elementary, in the sense that it relies essentially on two classical exponential inequalities (which are recalled in Appendix A). This proof is an adaptation of the proof of Proposition 6.7 in [2]. In Proposition 4.4 we shall then bound from below the probability of hitting a value larger than $m/\sqrt{\pi}$ before time $\kappa \ln m$ and this will conclude the proof of Theorem 2.2.

Proposition 4.3. *Let $\pi > 1$ be fixed. There exist $\rho > 1$, $c_0 > 0$, $m_0 \geq 1$, which depend on π and the ratio $f_0^*/\bar{f}_0$ only, such that, for any set of parameters ℓ, p_{C}, p_{M} satisfying $\pi = (f_0^*/\bar{f}_0)(1 - p_{\mathrm{C}})(1 - p_{\mathrm{M}})^{\ell}$, we have, for all $m \geq m_0$, $i \leq m/\sqrt{\pi}$,*

$$\mathbb{P}(N_{n+1} \leq \rho i \mid N_n = i) \leq \exp(-c_0 i).$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Proof. We recall that, 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 B_n is distributed according to the binomial law $\mathcal{B}(m/2, 1-p_C)$, the variables Y_k^i, $k \in \mathbb{N}$, $i \in \{1, \dots, m\}$, are Bernoulli random variables with parameter $\varepsilon_m(i)$, and all these random variables are independent. 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 large enough m, we have

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

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

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

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

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

Recall that $1 - p_C > \bar{f}_0/f_0^*$. For large enough m, we have

$$\frac{m}{2}(1-p_C) - l(m, \varepsilon) \ge \frac{m}{2}(1-p_C)\frac{\varepsilon}{2} - 1 \ge \frac{m\varepsilon}{4}\frac{\bar{f}_0}{f_0^*} - 1 \ge \frac{m\varepsilon}{8}\frac{\bar{f}_0}{f_0^*}.$$

It follows that, for large enough m,

$$\mathbb{P}(B_n \le l(m, \varepsilon)) \le \exp\left(-\frac{m}{32}\left(\frac{\varepsilon \bar{f}_0}{f_0^*}\right)^2\right).$$

Let us try to also apply Hoeffding's inequality to control the second probability. We obtain

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

Now

$$2l(m, \varepsilon)\varepsilon_m(i) \ge 2\frac{m}{2}(1-p_C)(1-\varepsilon)\frac{i f_0^*}{m\sqrt{\pi}\,\bar{f}_0}(1-p_M)^\ell = (1-\varepsilon)i\sqrt{\pi},$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

whence, using the hypothesis on ρ,

$$\mathbb{P}\biggl(\sum_{k=1}^{2l(m,\varepsilon)} Y_k^i < \rho i\biggr) \le \exp\biggl(-\frac{\pi\varepsilon^2 i^2}{m}\biggr).$$

This inequality becomes useful only when i is of the order δm 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 Chebyshev 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

$$b = \biggl\lfloor \frac{2l(m,\varepsilon) - (m/4)(1-p_{\mathrm{C}})\varepsilon}{i} + 1 \biggr\rfloor,$$

and we define the sum associated to each block of size b as, for all $j \in \{1,\dots,i\}$,

$$Y'_j = \sum_{k=b(j-1)+1}^{bj} Y_k^i.$$

Note that Y'_1 follows the binomial law with parameters b and $\varepsilon_m(i)$. We will next estimate from below the product $b\varepsilon_m(i)$. By the choice of b and l, we have

$$b \ge \frac{1}{i}\biggl(2l(m,\varepsilon) - \frac{m}{4}(1-p_{\mathrm{C}})\varepsilon\biggr), \qquad l(m,\varepsilon) \ge \frac{m}{2}(1-p_{\mathrm{C}})\biggl(1-\frac{\varepsilon}{2}\biggr),$$

whence

$$b \ge \frac{m}{i}(1-p_{\mathrm{C}})(1-\varepsilon) \quad \text{and} \quad \mathbb{E}(Y'_1) = b\varepsilon_m(i) \ge \sqrt{\pi}(1-\varepsilon) > \rho.$$

Let $\delta_0 > 0$ be such that $\delta_0 < (1-p_{\mathrm{C}})\varepsilon/4$. Suppose that $i \le \delta_0 m$. We also have

$$bi \le 2l(m,\varepsilon) - \frac{m}{4}(1-p_{\mathrm{C}})\varepsilon + i \le 2l(m,\varepsilon) - \frac{m}{4}(1-p_{\mathrm{C}})\varepsilon + \delta_0 m \le 2l(m,\varepsilon).$$

Using the Chebyshev exponential inequality (see Appendix A), we have

$$\begin{aligned}
\mathbb{P}\biggl(\sum_{k=1}^{2l(m,\varepsilon)} Y_k^i \le \rho i\biggr) &\le \mathbb{P}\biggl(\sum_{k=1}^{bi} Y_k^i \le \rho i\biggr) \\
&\le \mathbb{P}\biggl(\sum_{j=1}^{i} Y'_j \le \rho i\biggr) \\
&\le \mathbb{P}\biggl(\sum_{j=1}^{i} -Y'_j \ge -\rho i\biggr) \\
&\le \exp(-i\,\Lambda^*_{-Y'_1}(-\rho)),
\end{aligned}$$

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 use Lemma A.4 to compare the Cramér transforms of $-Y'_1$ and $-Y''_1$. We have

$$\Lambda^*_{-Y'_1}(-\rho) \ge \Lambda^*_{-Y''_1}(-\rho) = \rho\ln\biggl(\frac{\rho}{b\varepsilon_m(i)}\biggr) - \rho + b\varepsilon_m(i).$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

The map

$$\lambda \mapsto \rho \ln\left(\frac{\rho}{\lambda}\right) - \rho + \lambda$$

is nondecreasing on $[\rho, +\infty]$, and $b\varepsilon_m(i) \geq \sqrt{\pi}(1-\varepsilon)$, thus,

$$\Lambda^*_{-Y_1''}(-\rho) \geq \rho \ln\left(\frac{\rho}{\sqrt{\pi}(1-\varepsilon)}\right) - \rho + \sqrt{\pi}(1-\varepsilon).$$

Let us denote by c_0 the right-hand quantity. Then c_0 is positive and it depends only on ρ, π, $f_0^*/\bar{f}_0$, and ε. Finally, we have, for large enough m, for all $i \in \{1, \ldots, \lfloor \delta_0 m \rfloor\}$,

$$\mathbb{P}\left(\sum_{k=1}^{2l(m,\varepsilon)} Y_k^i \leq \rho i\right) \leq \exp(-c_0 i),$$

whence

$$\mathbb{P}(N_{n+1} \leq \rho i \mid N_n = i\,) \leq \exp\left(-\frac{m}{32}\left(\frac{\varepsilon \bar{f}_0}{f_0^*}\right)^2\right) + \exp(-c_0 i).$$

For i such that $\delta_0 m \leq i < m/\sqrt{\pi}$, we had obtained

$$\mathbb{P}(N_{n+1} \leq \rho i \mid N_n = i) \leq \exp\left(-\frac{m}{32}\left(\frac{\varepsilon \bar{f}_0}{f_0^*}\right)^2\right) + \exp(-\pi\varepsilon^2\delta_0^2 m).$$

Let $\eta \in (0, 1)$ be small enough so that $\eta c_0 \leq \pi\varepsilon^2\delta_0^2$ and, for large enough m,

$$\exp\left(-\frac{m}{32}\left(\frac{\varepsilon \bar{f}_0}{f_0^*}\right)^2\right) \leq \exp\left(-\eta\frac{mc_0}{2}\right)\left(1 - \exp\left(-\eta\frac{c_0}{2}\right)\right).$$

For large enough m and $i \in \{1, \ldots, \lfloor \delta_0 m \rfloor\}$, we have

$$\begin{aligned}\mathbb{P}(N_{n+1} \leq \rho i \mid N_n = i) &\leq \exp\left(-\eta\frac{ic_0}{2}\right)\left(1 - \exp\left(-\eta\frac{c_0}{2}\right)\right) + \exp(-\eta i c_0)\\ &\leq \exp\left(-\eta\frac{ic_0}{2}\right).\end{aligned}$$

For large enough m and $\delta_0 m \leq i < m/\sqrt{\pi}$, we also have

$$\begin{aligned}\mathbb{P}(N_{n+1} \leq \rho i \mid N_n = i) &\leq \exp\left(-\eta\frac{mc_0}{2}\right)\left(1 - \exp\left(-\eta\frac{c_0}{2}\right)\right) + \exp(-\eta m c_0)\\ &\leq \exp\left(-\eta\frac{ic_0}{2}\right).\end{aligned}$$

These inequalities yield the claim of the proposition. □

We define $\tau^* = \inf\{n \geq 0 \colon N_n \geq m/\sqrt{\pi}\}$.

Proposition 4.4. *Let $\pi > 1$ be fixed. There exist $\kappa > 0$ and $p^* > 0$, which depend on π and the ratio $f_0^*/\bar{f}_0$ only, such that, for all $m \geq 1$,*

$$\mathbb{P}(\tau^* \leq \kappa \ln m \mid N_0 = 1) \geq p^*.$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Proof. Let us define

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

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

$$\mathbb{P}(\tau^* \leq \kappa \ln m \mid N_0 = 1) = \mathbb{P}(\tau^* \leq \kappa \ln m,\ \tau^* < \tau_0 \mid N_0 = 1).$$

It is annoying to work with a Markov chain which has an absorbing state, so we first eliminate this problem. We consider the modified Markov chain $(\widetilde{N}_n)_{n\geq 0}$ 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 $(\widetilde{N}_n)_{n\geq 0}$, which is irreducible. Let $\rho > 1$, $c_0 > 0$, and $m_0 \geq 1$ be as given in Proposition 4.3. For $k \geq 0$, let T_k be the first time the process $(\widetilde{N}_n)_{n\geq 0}$ hits k, i.e.

$$T_k = \inf\{n \geq 0 \colon \widetilde{N}_n = k\}.$$

Let $\mathcal{E}$ be the event

$$\mathcal{E} = \left\{\text{for all } k < \frac{m}{\sqrt{\pi}},\ \widetilde{N}_{T_k+1} \geq \rho k\right\}.$$

We claim that, on the event $\mathcal{E}$, we have, for all $n \leq \tau^*$,

$$\widetilde{N}_{n+1} \geq \rho \widetilde{N}_n.$$

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

$$\widetilde{N}_{k+1} \geq \rho \widetilde{N}_k.$$

This implies, in particular, that

$$\widetilde{N}_0 < \widetilde{N}_1 < \cdots < \widetilde{N}_n < \frac{m}{\sqrt{\pi}}.$$

Suppose that $\widetilde{N}_n = i$. The above inequalities imply that $T_i = n$ and $\widetilde{N}_{T_i+1} = \widetilde{N}_{n+1} \geq \rho \widetilde{N}_n$, so that the inequality still holds at rank $n+1$. Iterating the inequality until time $\tau^* - 1$, we see that $\widetilde{N}_{\tau^*-1} \geq \rho^{\tau^*-1}$. Moreover, $\widetilde{N}_{\tau^*-1} \leq m/\sqrt{\pi}$, thus,

$$\tau^* \leq 1 + \frac{\ln m}{\ln \rho}.$$

Let $m_1 \geq 1$ and $\kappa > 0$ be such that, for all $m \geq m_1$,

$$1 + \frac{\ln m}{\ln \rho} \leq \kappa \ln m.$$

The constants m_1 and κ depend only on ρ, and we have

$$\mathbb{P}(\tau^* \leq \kappa \ln m, \tau^* < \tau_0 \mid \widetilde{N}_0 = 1) \geq \mathbb{P}(\mathcal{E}).$$

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

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Lemma 4.1. *Let $k \in \{1, \dots, m\}$ and let $i_1, \dots, i_k$ be k distinct points of $\{1, \dots, m\}$. The random variables $\widetilde{N}_{T_{i_1}+1}, \dots, \widetilde{N}_{T_{i_k}+1}$ are independent.*

Proof. We prove this 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 $i_1, \dots, i_k$ be k distinct points of $\{1, \dots, m\}$. Let $j_1, \dots, j_k$ be k points of $\{1, \dots, m\}$. Let us set

$$T = \min\{T(i_l) \colon 1 \leq l \leq k\}.$$

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

$$\begin{aligned}
\mathbb{P}(\widetilde{N}_{T(i_1)+1} = j_1, \dots, \widetilde{N}_{T(i_k)+1} = j_k) &= \sum_{1 \leq l \leq k} \mathbb{P}(\widetilde{N}_{T(i_1)+1} = j_1, \dots, \widetilde{N}_{T(i_k)+1} = j_k, T = T(i_l)) \\
&= \sum_{1 \leq l \leq k} \mathbb{P}(\widetilde{N}_{T(i_1)+1} = j_1, \dots, \widetilde{N}_{T(i_k)+1} = j_k \mid T = T(i_l)) \mathbb{P}(T = T(i_l)) \\
&= \sum_{1 \leq l \leq k} \mathbb{P}(\text{for all } h \neq l,\ \widetilde{N}_{T(i_h)+1} = j_h,\ \widetilde{N}_1 = j_l \mid \widetilde{N}_0 = i_l) \mathbb{P}(T = T(i_l)) \\
&= \sum_{1 \leq l \leq k} p(i_l, j_l) \mathbb{P}(\text{for all } h \neq l,\ \widetilde{N}_{T(i_h)+1} = j_h \mid \widetilde{N}_0 = j_l) \mathbb{P}(T = T(i_l)).
\end{aligned}$$

We use the induction hypothesis

$$\mathbb{P}(\text{for all } h \neq l,\ \widetilde{N}_{T(i_h)+1} = j_h \mid \widetilde{N}_0 = j_l) = \prod_{h \neq l} p(i_h, j_h).$$

Reporting in the sum, we obtain

$$\begin{aligned}
\mathbb{P}(\widetilde{N}_{T(i_1)+1} = j_1, \dots, \widetilde{N}_{T(i_k)+1} = j_k) &= \sum_{1 \leq l \leq k} \prod_{1 \leq h \leq k} p(i_h, j_h) \mathbb{P}(T = T(i_l)) \\
&= \prod_{1 \leq h \leq k} p(i_h, j_h).
\end{aligned}$$

This completes the induction step and the proof. □

We return to the proof of Proposition 4.4. Using Lemma 4.1 and Proposition 4.3, we obtain, for m larger than m_0 and m_1,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &\geq \prod_{1 \leq k \leq m} \mathbb{P}(\widetilde{N}_{T_k+1} \geq \rho k) \\
&= \prod_{1 \leq k \leq m} (1 - \mathbb{P}(N_1 < \rho k \mid N_0 = k)) \\
&\geq \prod_{1 \leq k \leq m} (1 - \exp(-c_0 k)) \\
&\geq \prod_{k=1}^{\infty} (1 - \exp(-c_0 k)).
\end{aligned}$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

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

$$p_2 = \min\{\mathbb{P}(\tau^* \le \kappa \ln m \mid N_0 = 1) \colon m \le \max(m_0, m_1)\}.$$

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

Appendix A

Monotonicity. We first recall some standard definitions concerning monotonicity and coupling for stochastic processes. A classical reference is [8], especially for applications to particle systems. In the next two definitions we consider a discrete-time Markov chain $(X_n)_{n\ge 0}$ with values in a space $\mathcal{E}$. We suppose that the state space $\mathcal{E}$ is finite and that it is equipped with a partial order $\le$. A function $f : \mathcal{E} \to \mathbb{R}$ is nondecreasing if, for all $x, y \in \mathcal{E}$,

$$x \le y \quad \Longrightarrow \quad f(x) \le f(y).$$

Definition A.1. The Markov chain $(X_n)_{n\ge 0}$ is said to be monotone if, for any nondecreasing function f, the function

$$x \in \mathcal{E} \mapsto \mathbb{E}(f(X_n) \mid X_0 = x)$$

is nondecreasing.

A natural way to prove monotonicity is to construct an adequate coupling. A coupling for the Markov chain $(X_n)_{n\ge 0}$ is a family of processes $(X_n^x)_{n\ge 0}$ indexed by $x \in \mathcal{E}$, which are all defined on the same probability space, and such that, for $x \in \mathcal{E}$, the process $(X_n^x)_{n\ge 0}$ is the Markov chain $(X_n)_{n\ge 0}$ starting from $X_0 = x$. The coupling is said to be monotone if, for all x, $y \in \mathcal{E}$,

$$x \le y \quad \Rightarrow \quad \text{for all } n \ge 1, \qquad X_n^x \le X_n^y.$$

If there exists a monotone coupling then the Markov chain is monotone.

Stochastic domination. Let μ and ν be two probability measures on $\mathbb{R}$. We say that ν stochastically dominates μ, which we denote by $\mu \preceq \nu$, if for any nondecreasing positive function f, we have $\mu(f) \le \nu(f)$.

Lemma A.1. *If μ and ν are two probability measures on $\mathbb{N}$ then μ is stochastically dominated by ν if and only if, for all $i \in \mathbb{N}$,*

$$\mu([i, +\infty)) \le \nu([i, +\infty)).$$

Proof. Let $f : \mathbb{N} \to \mathbb{R}^+$ be a nondecreasing function. We compute

$$\begin{aligned}\mu(f) &= \sum_{i\ge 0} \mu(i) f(i) \\ &= \sum_{i\ge 0} (\mu([i, +\infty)) - \mu([i+1, +\infty))) f(i) \\ &= f(0) + \sum_{i\ge 1} \mu([i, +\infty))(f(i) - f(i-1)).\end{aligned}$$

Under the above hypothesis, we conclude that indeed $\mu(f) \le \nu(f)$. □

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Lemma A.2. *Let $n \geq 1$, $p \in [0,1]$, $\lambda > 0$ be such that $(1-p)^n \geq \exp(-\lambda)$. Then the binomial law $\mathcal{B}(n,p)$ of parameters n and p is stochastically dominated by the Poisson law $\mathcal{P}(\lambda)$ of parameter λ.*

Proof. Let $X_1, \ldots, X_n$ be independent random variables with common law being Poisson of parameter $-\ln(1-p)$. Let Y be a further random variable, independent of $X_1, \ldots, X_n$, with law being Poisson of parameter $\lambda - n\ln(1-p)$. Obviously, we have

$$Y + X_1 + \cdots + X_n \geq \min(X_1, 1) + \cdots + \min(X_n, 1).$$

Moreover, the law of the left-hand side is the Poisson law of parameter λ, while the law of the right-hand side is the binomial law $\mathcal{B}(n,p)$. □

Lemma A.3. *Let $\lambda > 0$ and let Y be a random variable with law being Poisson $\mathcal{P}(\lambda)$ of parameter λ. For any $t \geq \lambda$, we have*

$$\mathbb{P}(Y \geq t) \leq \left(\frac{\lambda \mathrm{e}}{t}\right)^t.$$

Proof. We write

$$\begin{aligned}
\mathbb{P}(Y \geq t) &= \sum_{k \geq t} \frac{\lambda^k}{k!} \exp(-\lambda) \\
&= \sum_{k \geq t} \frac{\lambda^{k-t}}{k!} \exp(-\lambda)\lambda^t \\
&\leq \sum_{k \geq t} \frac{t^{k-t}}{k!} \exp(-\lambda)\lambda^t \\
&\leq \left(\frac{\lambda \mathrm{e}}{t}\right)^t.
\end{aligned}$$

□

Let Y be a random variable following the Poisson law $\mathcal{P}(\lambda)$. For any $t \in \mathbb{R}$, we have

$$\Lambda_Y(t) = \ln \mathbb{E}(\exp(tY)) = \ln\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \exp(-\lambda + kt)\right) = \lambda(\exp(t) - 1).$$

For any $\alpha, t \in \mathbb{R}$,

$$\Lambda_{\alpha Y}(t) = \Lambda_Y(\alpha t) = \lambda(\exp(\alpha t) - 1).$$

Let us compute the Fenchel–Legendre transform $\Lambda^*_{\alpha Y}$. By definition, for $x \in \mathbb{R}$,

$$\Lambda^*_{\alpha Y}(x) = \sup_{t \in \mathbb{R}} (tx - \lambda(\exp(\alpha t) - 1)).$$

The maximum is attained at $t = (1/\alpha)\ln(x/(\lambda\alpha))$; hence,

$$\Lambda^*_{\alpha Y}(x) = \frac{x}{\alpha} \ln\left(\frac{x}{\lambda\alpha}\right) - \frac{x}{\alpha} + \lambda.$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Lemma A.4. *Let $p \in [0,1]$ and let $n \geq 1$. Let X be a random variable following the binomial law $\mathcal{B}(n,p)$. Let Y be a random variable following the Poisson law $\mathcal{P}(np)$. For any $\alpha \in \mathbb{R}$, we have $\Lambda^*_{\alpha X} \geq \Lambda^*_{\alpha Y}$.*

Proof. For any $t \in \mathbb{R}$, we have

$$\Lambda_X(t) = \ln \mathbb{E}(\exp(tX)) = n \ln(1 - p + p\exp(t)) \leq np(\exp(t) - 1).$$

For any $\alpha, t \in \mathbb{R}$,

$$\Lambda_{\alpha X}(t) = \Lambda_X(\alpha t) \leq np(\exp(\alpha t) - 1).$$

We recall that, if Y is distributed according to the Poisson law of parameter λ, then, for all $t \in \mathbb{R}$,

$$\Lambda_Y(t) = \lambda(\exp(t) - 1).$$

Thus, taking $\lambda = np$, we conclude that, for all $t \in \mathbb{R}$,

$$\Lambda_{\alpha X}(t) \leq \Lambda_{\alpha Y}(t).$$

Taking the Fenchel–Legendre transform, we obtain, for all $x \in \mathbb{R}$,

$$\Lambda^*_{\alpha X}(x) \geq \Lambda^*_{\alpha Y}(x),$$

as required. □

Hoeffding's inequality. We state Hoeffding's inequality for Bernoulli random variables [6]. Suppose that X is a random variable with law being binomial $\mathcal{B}(n,p)$. We have, for all $t < np$,

$$\mathbb{P}(X < t) \leq \exp\left(-\frac{2}{n}(np - t)^2\right).$$

Chebyshev's exponential inequality. Let $X_1, \ldots, X_n$ be i.i.d. random variables with common law μ. Let Λ be the log-Laplace of μ, defined by, for all $t \in \mathbb{R}$,

$$\Lambda(t) = \ln\left(\int_{\mathbb{R}} \exp(ts)\,\mathrm{d}\mu(s)\right).$$

Let Λ^* be the Cramér transform of μ, defined by, for all $x \in \mathbb{R}$,

$$\Lambda^*(x) = \sup_{t \in \mathbb{R}} (tx - \Lambda(t)).$$

We suppose that μ is integrable and we denote by m its mean, i.e. $m = \int_{\mathbb{R}} x\,\mathrm{d}\mu(x)$. We then have (see, for instance, [3]), for all $x \geq m$,

$$\mathbb{P}\left(\frac{1}{n}(X_1 + \cdots + X_n) \geq x\right) \leq \exp(-n\Lambda^*(x)).$$

Let Y be a random variable following the Poisson law $\mathcal{P}(\lambda)$. For any $t \in \mathbb{R}$, we have

$$\Lambda_Y(t) = \ln \mathbb{E}(\exp(tY)) = \ln\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \exp(-\lambda + kt)\right) = \lambda(\exp(t) - 1).$$

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

For any $\alpha, t \in \mathbb{R}$,

$$\Lambda_{\alpha Y}(t) = \Lambda_Y(\alpha t) = \lambda(\exp(\alpha t) - 1).$$

Let us compute the Fenchel–Legendre transform $\Lambda^*_{\alpha Y}$. By definition, for $x \in \mathbb{R}$,

$$\Lambda^*_{\alpha Y}(x) = \sup_{t \in \mathbb{R}} (tx - \lambda(\exp(\alpha t) - 1)).$$

The maximum is attained at $t = (1/\alpha)\ln(x/(\lambda\alpha))$, hence,

$$\Lambda^*_{\alpha Y}(x) = \frac{x}{\alpha}\ln\left(\frac{x}{\lambda\alpha}\right) - \frac{x}{\alpha} + \lambda.$$

Galton–Watson processes. Let ν be a probability distribution on the nonnegative integers. Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables distributed according to ν. The Galton–Watson process with reproduction law ν is the sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ defined by $Z_0 = 1$ and, for all $n \in \mathbb{N}$,

$$Z_{n+1} = \sum_{k=1}^{Z_n} Y_k.$$

It is said to be subcritical if $\mathbb{E}(\nu) < 1$ and supercritical if $\mathbb{E}(\nu) > 1$. The following estimates are classical (see, for instance, [1]).

Lemma A.5. *Let $(Z_n)_{n\in\mathbb{N}}$ be a subcritical Galton–Watson process. There exists a positive constant c, which depends only on the law ν, such that, for all $n \geq 1$,*

$$\mathbb{P}(Z_n > 0) \leq \exp(-cn).$$

Proposition A.1. *Let $(Z_n)_{n\in\mathbb{N}}$ be a supercritical Galton–Watson process such that $\mathbb{E}(\nu)$ is finite. Let*

$$\tau_1 = \inf\{n \geq 1 \colon Z_n > n^{1/4}\}.$$

There exist $\kappa > 0$, $c_1 > 0$, $n_1 \geq 1$, such that, for all $n \geq n_1$,

$$\mathbb{P}(\tau_1 < \kappa \ln n) \leq \frac{1}{n^{c_1}}.$$

Proof. We have, for $k \geq 0$,

$$\begin{aligned}\mathbb{P}(\tau_1 = k) &\leq \mathbb{P}(\tau_1 \geq k, Z_k > n^{1/4}) \\ &\leq \mathbb{P}(Z_k > n^{1/4}) \\ &\leq n^{-1/4}\mathbb{E}(Z_k) \\ &\leq n^{-1/4}(\mathbb{E}(\nu))^k.\end{aligned}$$

We sum this inequality. For $n \geq 1$,

$$\mathbb{P}(\tau_1 < n) \leq n^{-1/4}\sum_{k=0}^{n-1}(\mathbb{E}(\nu))^k = n^{-1/4}\frac{(\mathbb{E}(\nu))^n - 1}{\mathbb{E}(\nu) - 1}.$$

We choose κ positive and sufficiently small, we apply this inequality with $\kappa \ln n$ instead of n and we obtain the desired conclusion. □

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press

Acknowledgement

I thank an anonymous referee for his/her remarks, which helped to improve the presentation of the paper.

References

[1] Athreya, K. B. and Ney, P. E. (2004). *Branching Processes*. Dover, Mineola, NY.
[2] Cerf, R. (2014). The quasispecies regime for the simple genetic algorithm with ranking selection. Preprint. Available at https://arxiv.org/abs/1403.5427.
[3] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd edn. Springer, New York.
[4] Eigen, M., McCaskill, J. and Schuster, P. (1989). The molecular quasi-species. In *Advances in Chemical Physics*, Vol. 75. John Wiley, Hoboken, NJ, pp. 149–263.
[5] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA.
[6] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58,** 13–30.
[7] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
[8] Liggett, T. M. (2005). *Interacting Particle Systems*. Springer, Berlin.
[9] Mills, K. L., Filliben, J. J. and Haines, A. L. (2015). Determining relative importance and effective settings for genetic algorithm control parameters. *Evolutionary Comput.* **23,** 309–342.

https://doi.org/10.1017/apr.2017.26 Published online by Cambridge University Press