Rate of convergence of crossover operators

Yann Ollivier
Laboratoire de Mathématique d'Orsay
UMR 8628 du CNRS
Bât. 425
Université de Paris–Sud
91405 Orsay, France
e-mail: Yann.Ollivier@math.u-psud.fr

ABSTRACT

We study the convergence of mating operators on $\{0,1\}^n$. In particular, we answer questions of Rabani, Rabinovich and Sinclair (*cf.* [5]) by giving tight estimates on the divergence between the finite- and infinite-population processes, thus solving positively the problem of the simulability of such quadratic dynamical systems.

Keywords: quadratic dynamical systems, crossover, mixing time

INTRODUCTION, MAIN RESULTS

We study from a theoretical point of view the rate of convergence of a mating operator between two "genomes", in the framework of population genetics or genetic algorithms: a population is made up of individuals defined by a genome, which is a string of symbols (taken in $\{0,1\}$ for convenience).

The mating operator consists in having a stem population replaced by a new one in the following way: an individual from the new population is obtained by randomly, uniformly sampling two distinct individuals in the stem population and mixing their genomes in some prescribed way. These operations are repeated independently in order to obtain all the individuals of the new population.

Intuitively, mating seems to mix the genes present in the stem population. Biology handbooks claim that the interest of sexual reproduction is to keep a high level of diversity and to mix all available genes. Thus, it can be interesting to study the speed of such a mixing.

We choose a genome length n, and we define the random offspring of a mating between two elements of $\{0,1\}^n$ as follows: Fix a probability distribution Π (a *crossover operator*) on the set of subsets of $\{1 \dots n\}$. Sample an $S \subset \{1 \dots n\}$ from Π. Then, the offspring of the pair $x, y \in \{0,1\}^n$ is a random element $z \in \{0,1\}^n$ whose i-th bit z_i is equal to x_i if $i \in S$, or y_i if $i \notin S$.

According to the chosen distribution Π, different kinds of mating can be obtained. The simplest one is *uniform crossover*: Π is the uniform distribution on all subsets of $\{1 \dots n\}$. This amounts to choosing each of the bits z_i to be equal to x_i or y_i independently of each other with probability $1/2$.

We consider a finite population process of size k: at any step, the population is made up of k (not necessarily distinct) elements of $\{0,1\}^n$. The population for the next step is obtained by uniformly picking, k times with replacement, a random pair of distinct individuals[1] in the previous population, by having them generate a child from our mating operator and by putting the child in the new population.

Let π_t be the random k-tuple in $\{0,1\}^n$ obtained after t iterations of the process, given an initial k-tuple π_0.

We want to compare this process with the so-called "infinite-population process" where an infinite population is a probability distribution on $\{0,1\}^n$: the law of an element from the distribution p_{t+1} is obtained by sampling two individuals according to p_t and mating them according to Π. For a given p_0, we obtain a (deterministic) sequence p_t of probability distributions on $\{0,1\}^n$.

The infinite-population process is fairly well-known (see the work by Y. Rabani, Y. Rabinovich and A. Sinclair in [5]). It converges to a distribution p_∞ which depends on p_0 in the following way: under p_∞, the bits of an individual are chosen independently of each other, and their value is 0 or 1 with the same probability as in p_0. In other words, the proportion, in the population, of 0 and 1 at each position in the genome is invariant under the process, but the values at different positions tend to be independent.

The authors of [5] give essentially tight upper and lower bounds on the convergence of the infinite-population process. Let us recall their main result.

Definition. *Let the distance $|p - p'|$ between two probability distributions p and p' be*

$$|p - p'| = \frac{1}{2} \sum_{x \in \{0,1\}^n} |p(x) - p'(x)| = \sup_{X \subset \{0,1\}^n} |p(X) - p'(X)| \leqslant 1$$

[1] By "distinct" individuals we do not mean that their genomes are necessarily distinct, but that they correspond to distinct indices $i, j \leqslant k$ in the population. This assumption is natural for the modelling of sexual reproduction in biological systems (excluding occasional parthenogenesis). If we release this assumption, all results stated here remain true with the constant r_Π replaced with $r_\Pi(1 - 1/k) + 1/k$.

Then, under a natural non-degeneracy assumption on the mating operator, we have $|p_t - p_\infty| \leqslant n^2\, r_\Pi^t$, where $r_\Pi < 1$ is a constant depending on the crossover operator (equal to $1/2$ for uniform crossover). Furthermore, these authors show that for particular crossover operators, this result is essentially tight, in the sense that e.g. for uniform crossover, the time required for $|p_t - p_\infty|$ to be less than $1/4$ (the "mixing time") is at least $\log_2 n - O(1)$ for some initial population. At the end of the paper we prove a similar but different tightness result (see section 2.A).

On the other hand, the finite-population process is harder to comprehend. It can be thought of as an approximation of the infinite-population process; but it seems that, in order to determine an individual at some step, it would be necessary to know its two parents, its four grandparents, ... its 2^t forefathers. Thus if the population is small, some forefathers will appear several times in the family tree, which will result in undesired correlations.

This problem arises for all so-called "quadratic dynamical systems" (*cf.* [6]), when we are given some random "mating" between two individuals in a given space, and we evolve probability measures on this space by defining the law of an individual at time $t+1$ to be the law of the offspring of two individuals picked from the law at time t. The difficulty of simulating a quadratic dynamical system has been formalized (*cf.* [1]): indeed, such systems can solve in polynomial time any PSpace problem.

The comparison between the two processes goes as follows: Given an infinite population p_0, we sample k individuals from it. This results in a random k-tuple π_0. This k-tuple evolves as described above, and we denote by π_t the k-tuple at time t.

Actually, π_t seen as a probability measure on $\{0,1\}^n$ (each element of the k-tuple having weight $1/k$) is of course not a good approximation of the infinite population p_t since it is supported on only k individuals, whereas in general p_t is supported on all of $\{0,1\}^n$ more or less uniformly.

We could rather try to compare the law of the random k-tuple π_t with the law $p_t^{\otimes k}$ of a random k-sample from p_t (after all, π_0 was a k-sample from p_0). As it turns out, this is not a good comparison. Indeed, after some time, π_t is very probably made up of k clones of one single individual (this is because at each step, with small probability, some genetic information gets lost). This well-known phenomenon is termed *coalescence*. (By the way, this shows that the process π_t converges.) We will return to this in section 3..

But the random individual making up this uniform population π_t will not always be the same, and its probability law will be close to p_t, which is what we wish. Thus, the law of a single element (e.g. the first one) of π_t, taken alone, is a good approximation to p_t.

Hence, denote by q_t the probability law of the first element of the random k-tuple π_t.

Y. Rabani, Y. Rabinovich and A. Sinclair prove that $|q_t - p_t| \leqslant \dfrac{4n^2 t}{k}$. Our main result is that

$$|q_t - p_\infty| \leqslant \frac{n^2}{C_\Pi\, k} + n^2\, r_\Pi^t$$

where r_Π is the same constant depending on the crossover operator as in Y. Rabani, Y. Rabinovich and A. Sinclair's result on the infinite-population process, and $C_\Pi = 1 - r_\Pi + 1/k$.

For example, for uniform crossover, this leads to

$$|q_t - p_\infty| \leqslant n^2 \left(\frac{2}{k+2} + \frac{1}{2^t} \right)$$

Considering uniform crossover, letting $k \to \infty$ so that the finite-population process closely follows the infinite-population process, and applying our lower bound stated above in that case, shows that the term $n^2/2^t$ (with $r_\Pi = 1/2$) is tight up to a $O(n)$ factor.

Furthermore, we prove that for k big enough, for some initial population, we have $|p_\infty - q_\infty| \geqslant \dfrac{n}{Ck}$ for some constant $C \leqslant 32$. So, our bounds are essentially tight up to replacement of n^2 by n, which affect the mixing time by at most a factor of 2.

At the end of the paper, we give a proof of similar results regarding mean-time (before coalescence) approximation of a whole population rather than a single individual.

1. CONVERGENCE OF THE FINITE POPULATION PROCESS

1.A Background : convergence of the infinite population process

We recall here the results of Y. Rabani, Y. Rabinovich and A. Sinclair (*cf.* [5]).

Let p_0 be a probability distribution on $\{0,1\}^n$. Let a_{i0} be the probability that the i-th bit of an individual sampled from p_0 is 0, and $a_{i1} = 1 - a_{i0}$.

Denote by p_∞ the probability law which, to the individual $x = x_1 x_2 \dots x_n$, assigns the weight $p_\infty(x_1 x_2 \dots x_n) = \prod a_{ix_i}$. This is the probability law where each bit equals 0 or 1 with the same probability as in p_0, but where different bits are chosen independently of each other.

For example, if p_0 is the distribution that puts weight 1/2 on the individual $000\dots0$ and 1/2 on $111\dots1$, then p_∞ is the uniform distribution on $\{0,1\}^n$.

Here, Y. Rabani, Y. Rabinovich and A. Sinclair make a non-degeneracy assumption on the chosen crossover operator Π: they demand that each two different positions $1 \leqslant i, j \leqslant n$ have a positive probability to be separated by the crossover, that is, that there be an $S \subset \{1 \dots n\}$ with $\Pi(S) > 0$ and $i \in S, j \notin S$ (otherwise, these two positions could be considered as one single two-bit block).

This natural assumption holds for all usual crossovers. The authors are especially interested in the following cases:

- Uniform crossover: Π is the uniform distribution on subsets of $\{1 \dots n\}$, each bit is picked independently from one of the two parents.
- One-point crossover: Choose a position $1 \leqslant i \leqslant n+1$ uniformly. Those bits with position less than i will be picked from one parent and the other bits from the other one. So Π gives equal weight to the $n+1$ sets $\varnothing$, $\{1\}$, $\{1,2\}$, $\{1,2,3\}$, $\dots$, $\{1,2,\dots,n\}$.

- Poisson crossover: We begin at position 0, picking successive bits of one parent. Then after some time we jump to the other parent and pick some successive bits from it, etc. At each step, the probability to jump from one parent to the other is the same.

Under the non-degeneracy assumption, [5] states the following result:

Theorem 1.1 [5]. *The infinite-population process p_t converges to p_∞ (as probability measures on $\{0,1\}^n$).*

Furthermore, they give a good estimate of the rate of convergence. This depends on details of the crossover operator. Following their notation, let $r_{ij}(\Pi)$ be the probability that positions i and j are *not* separated by an $S \subset \{1 \dots n\}$ sampled from Π. Let $r_\Pi = \max_{i,j} r_{ij}(\Pi)$. The non-degeneracy assumption states that $r_\Pi < 1$.

Then

Theorem 1.2 [5]. *The distance between the population at time t and the limit population p_∞ satisfies*

$$|p_t - p_\infty| \leqslant n^2\, r_\Pi^t$$

For instance, $r_\Pi = 1/2$ for uniform crossover, and hence $|p_t - p_\infty| \leqslant n^2/2^t$.

1.B Convergence for finite populations

Recall that q_t is the law of the first element of the random k-tuple π_t after t steps of the finite-population process, when π_0 is made up of k independent samplings from p_0.

In [5], Y. Rabani, Y. Rabinovich and A. Sinclair show that

$$|q_t - p_t| \leqslant \frac{4n^2 t}{k}$$

for any crossover operator. We show here, using similar techniques, that

Theorem 1.3.

$$|q_t - p_\infty| \leqslant \frac{n^2}{k\,(1 - r_\Pi + 1/k)} + n^2\, r_\Pi^t$$

with r_Π as above.

In particular, $|q_\infty - p_\infty| \leqslant n^2/(k(1 - r_\Pi + 1/k))$, and for $k \gg n^2$, the mixing time is less than $2 \log_{1/r_\Pi} n$.

For the sake of optimality, we show below (section 2.B) that $|q_\infty - p_\infty| \geqslant n/Ck$ in some cases (where C is a constant). This is an intrinsic bias due to the finite population approximation.

Note that this bound is *not* obtained from bounding $|q_t - p_t|$ and then using the bound on $|p_t - p_\infty|$. We directly compare q_t with p_∞. The bias of p_t and of q_t compared to p_∞ may not be of the same kind.

Proof. We look at the process π_t in the following way: To generate π_t, we first leave π_0 unspecified, we choose a family tree from generation 0 to generation t from the correct probability distribution, and, fully independently, we fill π_0 by sampling k individuals from p_0. Then we look at how the bits of generation 0 propagate through the tree.

More precisely, a "family tree" is a structure in which, for each $t \geqslant 1$ and for each individual number i in generation t, two distinct members i_1 and i_2 of the previous generation are specified, together with a "mask" $S \subset \{1 \dots n\}$ describing those bits of i that come from i_1 or i_2. This tree gets a probability, which is the product of the probabilities, under Π, of all masks appearing in it, divided by $(k(k-1))^{kt}$ which corresponds to all possible choices of the parents of all individuals.

Once a family tree is given, we fill the bits of generation 0 using the distribution p_0, independently of this tree. Under these conditions, we are in a position to travel back through the tree and tell, for each bit of any individual at generation t, which bit from which individual of generation 0 it comes from.

We then note that, if we get a tree such that all n bits of the first individual of generation t come from *distinct* individuals from generation 0, these n bits come from n individuals independently sampled from p_0. The values, 0 or 1, of these bits are thus independent, and the i-th bit is a 1 with probability a_{i1} (in our earlier notation). In other words, if we get a tree where the n bits of the first individual of π_t come from distinct individuals, then the law of this individual is exactly p_∞ and we are done.

Then, a little manipulation of the definition of $|q_t - p_\infty|$ shows that this distance is less than the probability that the sampled tree be not of the above kind.

Let's evaluate this probability. Consider two bits of the first individual at generation t. If at some time $t' \leqslant t$, these bits belong to the same individual, the probability that they come from the same parent of this individual at time $t'-1$ is a number p depending on Π, with $p \leqslant r_\Pi$. If at time t' they belong to two different individuals, their respective parents are chosen independently in $\pi_{t'-1}$, and the probability that they come from the same individual of $\pi_{t'-1}$ is $1/k$.

Going back through the tree, we thus have a Markov chain with the following transition probabilities between the two states D (the two bits belong to two distinct individuals) and S (they belong to the same individual): $D \to D$ with probability $1-1/k$, $D \to S$ with probability $1/k$, $S \to S$ with probability $p \leqslant r_\Pi$, $S \to D$ with probability $1-p$.

A (very simple) calculation gives that, knowing that at time t the bits are together, the probability to get a family tree where these two bits are together at time 0 is

$$\frac{1}{k(1-p+1/k)} + \left(p - \frac{1}{k}\right)^t \left(1 - \frac{1}{k(1-p+1/k)}\right)$$

which, since $p \leqslant r_\Pi$, is less than

$$\frac{1}{k(1-r_\Pi+1/k)} + \max(r_\Pi, 1/k)^t$$

In general, $1/k$ will be smaller than r_Π. If not, note that $\dfrac{1}{k(1-r_\Pi+1/k)}+\dfrac{1}{k^t} \leqslant \dfrac{2}{k(1-r_\Pi+1/k)}$ as soon as $k \geqslant 2, t \geqslant 2$ (the cases $k=1$ or $t=0,1$ being trivial). Anyway, the probability in question is less than $\dfrac{2}{k(1-r_\Pi+1/k)}+r_\Pi^t$.

This was for one pair of bits of the first individual of generation t. There are $n(n-1)/2$ such pairs. The probability that the sampled tree presents two bits with the same ancestor from generation 0 is, then, less than $\dfrac{n(n-1)}{2}\left(\dfrac{2}{k(1-r_\Pi+1/k)}+r_\Pi^t\right)$, hence the theorem. ∎

The main difference with the analysis in [5] is that we make a more refined analysis of collisions: collisions are not so much disturbing, as two bits which collide at some time can be separated again further back in the tree. Note that this leads to a comparison of q_t to p_∞ and not to p_t, because once a collision has occurred the correlation between q_t and p_t is lost, and further separation of the collided bits does not restore this correlation which relies on the specific structure of the tree.

2. LOWER BOUNDS ON CONVERGENCE

We now turn to proving that the bounds for convergence obtained so far are essentially tight. Results in this direction for infinite populations already appear in [5]. We give below a tightness result for finite populations. As a template, we begin by giving a tightness result for uniform crossover in infinite populations which is different from that of [5].

2.A Lower bound for uniform crossover in infinite populations

Recall Theorem 1.2: $|p_t - p_\infty| \leqslant n^2\, r_\Pi^t$. The asymptotic part (in t) of this is tight: indeed, there exists a population p_0 such that for all t, $|p_t - p_\infty| \geqslant r_\Pi^t/2$.

Define the mixing time τ of the process as the smallest t such that whatever the initial population p_0 was, we have $|p_t - p_\infty| < 1/4$. So $\tau \leqslant 2\log_{1/r_\Pi} n + 2\log_{1/r_\Pi} 2$, which is a fairly good result.

The authors of [5] show that for particular crossover operators, this result is essentially tight. Their argument depends on the details of the crossover. For instance, for uniform crossover, they obtain $\tau \geqslant \log_2 n - O(1)$; hence the bound on the mixing time is tight up to a factor of 2. For Poisson crossover, their result is tight up to a factor of $O(\log\log\log n)$.

We prove that for uniform crossover, the result is essentially tight in a different sense than that of [5]. Namely, we show that for some initial population p_0, we have $|p_t - p_\infty| \geqslant \dfrac{n}{C\,2^t}$ for some constant C, for t large enough. So we cannot replace n^2 by an expression smaller than n in the upper bound above for $|p_t - p_\infty|$.

These results are not directly comparable: the one deals with the time required to reach some threshold, whereas the other reflects the asymptotic behavior. However, assuming that our estimate of the asymptotic behavior is tight even for short times would result in the same estimate $\log_2 n - O(1)$ for the mixing time.

Theorem 2.1. *For uniform crossover, for n and t large enough, for some initial population p_0, we have*

$$|p_t - p_\infty| \geqslant \frac{n}{32 \cdot 2^t}$$

Inspecting the proof reveals that the result holds as soon as $n \geqslant 8$ and $t \geqslant 3 \log_2 n + 4$ (the time from which the theorem holds depends inevitably on n, since otherwise $n/(32\ 2^t)$ could be greater than 1).

Proof. Let us have a fresh look at how an individual from generation t is built. First, let's fix the 2^t ancestors of this individual at time 0, sampled from p_0. Then, we observe that, under uniform crossover, each of the n bits of the individual comes from one of these ancestors, which we will call the ancestor of the specified bit. In the case of uniform crossover, by a straightforward induction, the ancestor of each bit is chosen uniformly and independently among the 2^t ancestors of the given individual (this is specific to uniform crossover). In other words, the distribution of the ancestors of the n bits of an individual is an independent sampling with replacement of n individuals among its 2^t ancestors.

We will use the fact that, sometimes, two bits come from the same ancestor to evaluate the deviation of p_t from p_∞. For this purpose, we will take as our p_0 the distribution on $\{0,1\}^n$ putting weight $1/2$ on the individual $111\ldots1$ and $1/2$ on $000\ldots0$. We will consider the law of the number of 1's in an individual under p_∞ and p_t, and find a difference.

Under p_∞, the law of the number of 1's is binomial with parameters n and $1/2$.

Under p_t, each bit of an individual comes from one of its ancestors at time 0. If the n ancestors of the n bits are all distinct, then these bits are picked uniformly and independently from p_0, in which case we find again a binomial distribution.

If, conversely, two bits of an individual come from the same ancestor at time 0, given our population p_0, these two bits will be equal. This leads to correlations which result in a quantifiable difference in the law of the number of 1's in an individual.

We will first evaluate the deviation obtained when exactly two bits have the same ancestor. We will then show that exactly two bits have the same ancestor with a large enough probability, and that the cases when more than one correlation occurs have a negligible weight when t is large. The first statement is the subject of the following lemma.

Lemma 2.2. *Let $n \geqslant 8$. Let μ_1 be the uniform probability measure on $\{0,1\}^n$. Let μ_2 be the measure on $\{0,1\}^n$ equal to $1/2^{n-1}$ at those points $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ such that $x_1 = x_2$ and equal to 0 elsewhere. Then, the difference between the probabilities under μ_1 and μ_2 of the event "the number of 1's in a sample individual lies between $n/2 - \sqrt{n/8}$ and $n/2 + \sqrt{n/8}$" is larger than $1/2n$.*

Proof of the lemma. Under μ_1, without correlated bits, the law of the number of 1's is a binomial $\binom{n}{r}/2^n$.

Under μ_2, there is one pair of correlated bits, and the law of the number of 1's will rather be $\frac{1}{2}\binom{n-2}{r-2}/2^{n-2} + \frac{1}{2}\binom{n-2}{r}/2^{n-2}$ (respectively for the cases when the correlated pair is made up of two 1's or two 0's).

The latter is less than the former in a zone around $n/2$, and greater elsewhere. The difference between the two is $\left(\binom{n}{r} - 2\binom{n-2}{r-2} - 2\binom{n-2}{r}\right)/2^n$, which is, after a small calculation, equal to $\binom{n-2}{r-1}/2^n \left(\frac{n-(n-2r)^2}{2(n-r)r}\right)$. The second term is positive for $n/2 - \sqrt{n}/2 \leqslant r \leqslant n/2 + \sqrt{n}/2$, equal to $2/n$ at $r = n/2$; it is greater than $1/n$ for $|r - n/2| \leqslant \sqrt{n/8}$.

Then, the difference of the probabilities under μ_1 and μ_2 that the number of 1's falls between $n/2 - \sqrt{n/8}$ and $n/2 + \sqrt{n/8}$ is greater than

$$\frac{1}{n} \sum_{r=n/2-\sqrt{n/8}}^{n/2+\sqrt{n/8}} \frac{1}{2^n}\binom{n-2}{r-1}$$

Knowing that a binomial of parameter $1/2$ is almost a bell curve:

$$\sum_{r=n/2-\sqrt{n/8}}^{n/2+\sqrt{n/8}} \frac{1}{2^{n-2}}\binom{n-2}{r-1} \sim \frac{2}{\sqrt{\pi}} \int_{-1/\sqrt{2}}^{1/\sqrt{2}} e^{-x^2/2} \mathrm{d}x \geqslant 1/2$$

(and the first term is indeed greater than $1/2$ as soon as $n \geqslant 8$), we get that this expression is greater than $1/2n$, which proves the lemma. ∎

Observe that, if the n bits of an individual at step t have distinct ancestors at step 0, the law of the number of 1's in these n bits is the same as under μ_1 in the lemma. If exactly two bits have the same ancestor, the law of the number of 1's will be the same as under μ_2 in the lemma.

We will now derive from this an evaluation of the distance between p_t and p_∞. Let A_0 be the event "all n bits have distinct ancestors at time 0", A_1 the event "exactly one pair of bits has a common ancestor", A_2 the remaining cases (more than one coincidence). Let also B be the event "the number of 1's falls between $n/2 - \sqrt{n/8}$ and $n/2 + \sqrt{n/8}$".

Then, according to the lemma:

$$\begin{aligned}
|p_\infty - p_t| &\geqslant |p_\infty(B) - p_t(B)| \\
&\geqslant |(p_\infty(B) - p_t(B|A_0))\, p_t(A_0) + (p_\infty(B) - p_t(B|A_1))\, p_t(A_1)| \\
&\quad - |p_\infty(B) - p_t(B|A_2)|\, p_t(A_2) \\
&\geqslant 0 + \frac{1}{2n} p_t(A_1) - p_t(A_2)
\end{aligned}$$

(Knowing A_0, the number of 1's under p_t is the same as under p_∞.)

Thus, the issue is to evaluate the probabilities that exactly two bits, or more than two bits, have a common ancestor. We know that these ancestors are sampled

uniformly and independently from a set of size 2^t. We state the following lemma, which we will use again later.

Lemma 2.3. *If n (distinguishible) individuals are placed uniformly at random into k cells, the probability that exactly two elements are placed in the same cell is greater than* $\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)$.

Proof of the lemma. By elementary combinatorics, this probability is $\frac{1}{k^n}\frac{n(n-1)}{2}\,k(k-1)\dots(k-n+2)$, that is $\frac{n(n-1)}{2k}\,1\,(1-1/k)\dots(1-(n-2)/k)$, which is greater than $\frac{n^2}{4k}\left(1-\frac{n^2}{2k}\right)$. ∎

Let's denote $k=2^t$. By Lemma 2.3, the probability that exactly two bits of an individual have the same ancestor at time 0 is more than $n^2/8k$ for k large enough. The case when more than two correlations would occur, that is, at least two pairs of bits with common ancestors, or at least three bits with the same ancestor, has a probability not greater than n^4/k^2, which is of greater order in $1/k$. Indeed, the probability that two pairs have common ancestors is at most $(n(n-1)/2)^2\,k(k-1)k^{n-4}/k^n = O(n^4/k^2)$, and the probability that three bits have the same ancestor is $(n(n-1)(n-2)/6)\,k^{n-2} = O(n^3/k^2)$.

We saw above that $|p_\infty - p_t| \geqslant 1/(2n)\,p_t(A_1) - p_t(A_2)$. If we take $k \geqslant 16n^3$ we ensure that the probability of A_2 is less than $n/32k$, in which case the expression at play is no less than $n/32k$. ∎

2.B Lower bound for finite populations

It is instructive to note that the difference between the laws q_∞ and p_∞ cannot be interpreted as an error due to the sampling with replacement in the k-tuple π_0 of the genes of an individual of q_∞, as opposed to sampling without replacement in p_∞. Indeed, if that were the case, the probability that two genes of an individual of π_∞ come from the same individual in π_0 would be exactly $1/k$, whereas we have just seen that it is actually $1/(k(1-r_\Pi+1/k))$, which is greater especially for large populations.

The above analysis shows that $|q_\infty - p_\infty| \leqslant \frac{n^2}{k\,(1-r_\Pi+1/k)}$. Let's prove a corresponding lower bound, which shows we cannot improve this result by much:

Theorem 2.4. *For all $n \geqslant 2$, for k large enough, there exists some initial population such that*

$$|q_\infty - p_\infty| \geqslant \frac{n}{32k}$$

The k above which the proposition holds depends on n (otherwise, $n/32k$ could be more than 1). Since this is a negative asymptotic result, we will not worry too

much about an explicit value for the k above which the proposition holds; a crude inspection of the proof reveals it holds at least for $k \geqslant 48(2n)^{n^2+2}/(1-r_\Pi)^{n^2}$. Of course this is probably a gross overestimate.

Proof. As usual, we will consider an individual at time t, and look at the individuals at time 0 from which its n bits arise. We will have a close look at the distribution of these n individuals.

We will essentially work as in section 2.A: we will show that, with some probability of order n^2/k, exactly two bits have the same ancestor, which introduces a deviation of order $1/n$.

First, we will evaluate the probability that exactly two bits have a common ancestor at time 0. This probability is greater than the probability that exactly two bits have the same ancestor at time 0 and that, in addition, all bits are separated at time 1.

In the proof of theorem 1.3, we saw that the probability that some two bits of an individual of π_∞ have the same ancestor at time 1 is less than $\dfrac{n^2}{k(1-r_\Pi+1/k)}$. Thus, the probability that all of them are separated at time 1 is greater than $1-\dfrac{n^2}{k(1-r_\Pi+1/k)}$.

Now, if all bits are separated at time 1, their parents at time 0 are simply picked uniformly and independently among k. According to lemma 2.3, the probability that exactly two of them fall together is greater than $(n^2/4k)(1-n^2/2k)$.

Thus, the (unconditional) probability that at time 0, exactly two bits fall together is greater than $\dfrac{n^2}{4k}\left(1-\dfrac{n^2}{2k}\right)\left(1-\dfrac{n^2}{k(1-r_\Pi+1/k)}\right)$ which in turn is more than $\dfrac{n^2}{8k}$ as soon as k is large enough, say $k \geqslant 3n^2/(1-r_\Pi)$.

Under the assumption that there exist two bits with the same ancestor, we will find a deviation between the probabilities of some event under p_∞ and q_t. Of course, we will take as our p_0 the probability distribution on $\{0,1\}^n$ which puts weight $1/2$ on $111\ldots1$ and $1/2$ on $000\ldots0$. Then, we will be interested in the distribution of the number of 1's in an individual of generation t.

We will argue as in section 2.A. To do this, we must first establish that the case when exactly two bits of an individual at time t have the same ancestor at time 0 is predominant over the cases when there are more coincidences. This is the subject of the following lemma, which states that the distribution of the ancestors of the n bits of an individual has roughly the same asymptotics, when $k \to \infty$, as if these ancestors were sampled uniformly and independently among the k individuals of the initial population.

In particular, the cases when exactly two bits have the same ancestor will have a probability of order $1/k$, whereas those when more correlations occur will weigh for less than $1/k^2$. We measure the number of coincidences by the number of distinct individuals from which the n bits of our individual at time t come from. This lemma can be of independent interest.

Lemma 2.5. *There exist constants $C_{n,\Pi}$ and $C'_{n,\Pi}$ such that the probability that the n bits of an individual at time $t = \infty$ come from m distinct individuals from time 0 lies between $\frac{C_{n,\Pi}}{k^{n-m}}$ and $\frac{C'_{n,\Pi}}{k^{n-m}}$, for k large enough.*

(It is easy to see that it makes sense to speak about an individual from generation $t = \infty$: the process is Markovian on the space of k-individual populations. Often we will look at the process backwards, as if it started at $t = \infty$; this can easily be made rigorous by taking t large enough afterwards.)

Proof of the lemma. Let's fix an individual from generation t, $t \approx \infty$ (i.e. t large enough). We have already seen that for $n = m$, the probability that all its bits have distinct ancestors at time 0 is greater than $1 - O(1/k)$, for large t. (The constants implied in $O()$ depend of course on n and Π.)

The idea is to consider the Markov chain made up of the positions (in the k-individual population) of the ancestors of the n bits of the given individual, at time $t - t'$ (a Markov chain in t'). We will split this Markov chain into classes, the class m being made up of those situations when the n bits are distributed over $m \leqslant n$ individuals at time $t-t'$. We will consider the communication probabilities between these classes, and study the weight of these classes in equilibrium when t' tends to infinity (relative to t, but we take a large t).

Let $m(t')$ be the number of distinct individuals which the n bits come from at time $t - t'$, and $s(m)$ the probability that $m(\infty) = m$. We intend to show that $s(m) = O(1/k^{n-m})$. We already know that for $m < n - 1$, $s(m) = O(1/k)$. In the following, the constants implied by O depend on n, m and r_Π; we only intend to study the asymptotic behavior in k.

Now, let's estimate the distribution of $m(t' + 1)$ for a given $m(t')$.

To go from generation $t - t'$ to generation $t - t' - 1$, we consider the $m(t')$ individuals carrying the n bits. We decompose the process into two steps. In the first one, we consider the $m(t')$ blocks of bits, and we apply the mating operator Π to find $2m(t')$ "abstract parents" generating them. Among these $2m(t')$, only m', where $m(t') \leqslant m' \leqslant n$, carry some bits. In the second step, we paste back these m' abstract parents onto the population at time $t - t' - 1$, which is made up of k individuals. The pasting consists in choosing, for each of the m' abstract parents, which individual among the k it really is. These individuals are chosen independenlty and uniformly among k (there is some additional complication due to the fact that the two parents of one individual are distinct, in which case we choose among $k - 1$ rather than k, which does not affect the calculation much).

The probability that these m' parents are spread over $m'' \leqslant m'$ individuals of generation $t-t'-1$ is, by elementary combinatorics, of order $C_{m'}/k^{m'-m''}$ for large k. Now, knowing $m(t')$, we know that $m' \geqslant m(t')$ and that, moreover, if $m(t') < n$, then $m' > m(t')$ with probability greater than $1 - r_\Pi$.

In other words, the first of our two steps cannot decrease $m(t')$, and increases it with probability greater than $1-r_\Pi$ (if $m(t) < n$); the second one decreases the result with controlled probability, going from m' to m'' with probability $O(1/k^{m'-m''})$. All in all, $m(t'+1) < m(t')$ with probability $O(1/k^{m(t')-m(t'+1)})$, $m(t'+1) = m(t')$ with probability less than $r_\Pi + O(1/k)$, and $m(t' + 1) > m(t')$ otherwise: gener-

ally, the number of blocks of bits increases, and it decreases only with probabilities controlled by powers of k.

Let's move to the proof proper. We work by backwards induction on m.

Suppose we have already proved that for all $m' \leqslant m$, we have $s(m') = O(1/k^{n-m})$, and that for $m \leqslant m' \leqslant n$ we have $s(m') = O(1/k^{n-m'})$. Now, the probability $s(1)$ that at time 0 ($t' \approx \infty$), all bits lie together, is such that $s(1) \leqslant r_\Pi\, s(1) + O(1/k)\, s(2) + O(1/k^2)\, s(3) + \ldots + O(1/k^{n-1})\, s(n)$ (in equilibrium). According to our induction hypothesis, and since $r_\Pi < 1$, this is $O(1/k^{m+1})$.

Similarly, $s(2) \leqslant s(1) + r_\Pi\, s(2) + O(1/k)\, s(3) + \ldots + O(1/k^{n-2})\, s(n)$, which is $O(1/k^{m+1})$ by our induction hypothesis, and since $r_\Pi < 1$.

Step by step, up to $m' = m - 1$, we get that for $m' \leqslant m - 1$, we have $s(m) = O(1/k^{n-m+1})$, which concludes our induction and ends the proof of the upper bound in the lemma (the constants in the notation O depend on everything except k).

In order to get the lower bound in the lemma, it is enough to observe that $s(n) = 1 - O(1/k)$ and to note that the transition coefficients $n \to m$ from the state $m(t') = n$ to $m(t' - 1) = m$ are of order $1/k^{n-m}$. ∎

On one hand, we proved that exactly two bits have a common ancestor with probability greater than $n^2/8k$; on the other hand, the case when more than one pair of bits have a common ancestor has probability at most $O(1/k^2)$. It is then enough to take k large and apply lemma 2.2 to conclude. ∎

3. COALESCENCE AND MEAN-TIME APPROXIMATION OF A POPULATION

The results stated above deal with extraction of one individual from the finite population π_t. One can wonder if the law of the whole k-tuple π_t is close to, for example, the law $p_\infty^{\otimes k}$ of an independent k-sample from p_∞. This is false due to the coalescence phenomenon.

The following is a classical result in the so-called Wright-Fisher model (see e.g. [7], [3], [2] or [4]).

Proposition 3.1. *For large k, for all $\varepsilon > 0$, for*

$$t \geqslant 4k\,(\ln n - \ln \varepsilon + \ln 2)$$

then, with probability greater than $1 - \varepsilon$, the k-tuple π_t is made up of k copies of the same individual.

The k above which the proposition holds is independent of n and ε.

Corollary. Under the same assumptions, the distance $\left|\sigma_t - p_\infty^{\otimes k}\right|$ is greater than

$$1 - \varepsilon - \prod_{1 \leqslant i \leqslant n} \left(a_i^k + (1 - a_i)^k\right)$$

where σ_t is the law of the k-tuple π_t, which is a probability distribution on $(\{0,1\}^n)^k$.

Proof. Indeed, $\prod_{1\leqslant i\leqslant n}\left(a_i^k+(1-a_i)^k\right)$ is the weight, under $p_\infty^{\otimes k}$, of k-tuples made up of identical individuals. ∎

However, even for k not too large, the coalescence time $4k\log n$ is much larger than the characteristic time of the convergence $q_t\to p_\infty$, which is of order $2\log_{1/r_\Pi} n$. So hopefully, in the meantime, some number $m\leqslant k$ of individuals could be extracted from π_t, whose joint law would be close to $p_\infty^{\otimes m}$.

Indeed:

Theorem 3.2. *Let $m\leqslant k$. Let q_t^m be the joint law in $(\{0,1\}^n)^m$ of the first m individuals of π_t. Then*

$$\left|q_t^m-p_\infty^{\otimes m}\right|\leqslant\frac{m^2n^2}{k(1-r_\Pi+1/k)}+\frac{m^2n}{k}\,t+mn^2\,r_\Pi^t$$

Of course, "the m first individuals" could be replaced by any m-tuple chosen in advance among π_t.

The first term corresponds to the intrinsic bias of the finite population, even for long times, as studied above. The second reflects coalescence. The third renders the convergence to p_∞.

Note that k must be of order $(mn)^2$ for a non-trivial estimate.

The optimum in t (tradeoff between coalescence and convergence to p_∞) is achieved for $t\approx\log_{1/r_\Pi}\dfrac{nk}{m}$ and is roughly $\dfrac{m^2n^2}{k(1-r_\Pi+1/k)}+\dfrac{nm^2}{k}\log_{1/r_\Pi}\dfrac{nk}{m}$.

Using the same techniques as before (evaluating the number of 1's among the mn bits when two bits have the same ancestor), one may derive a lower bound, which matches the upper bound up to a factor of $1/mn$ (and constants), for large k and a given t.

Proof. We will follow the ancestry of the mn bits of the first m individuals of π_t. If these mn bits come from distinct individuals of π_0 (which requires $k\geqslant mn$), then the resulting distribution will be $p_\infty^{\otimes m}$.

Let us consider two given bits among these mn. If they are two different bits from the same individual, nothing changes in regard to our previous analysis, and the probability that they are not separated at time 0 is less than $\dfrac{1}{k(1-r_\Pi+1/k)}+r_\Pi^t$.

If these two bits are located at different positions in two different individuals of π_t, then the Markov chain describing their separation is the same. However, initially, they are separated. Their probability of falling together at a given time begins at 0 and tends geometrically to $\dfrac{1}{k(1-r_\Pi+1/k)}$; it is always less than $\dfrac{2}{k(1-r_\Pi+1/k)}$.

However, the picture is quite different if we consider two bits located at the same position in two individuals of π_t: indeed, if, somewhere in the family tree, these two bits are gathered into one single individual, they are actually the *same* bit, inherited from that individual. Going back further in the tree, up to π_0, the bits can never again be separated.

Given these two bits, at each (backward) generation, their gathering occurs when they have the same parent, i.e. with probability $1/k$. The probability of their

gathering in t backward steps is, thus, less than t/k (which is essentially tight for large k).

There are $mn(n-1)/2$ pairs of bits at different positions in a single individual; $m(m-1)n(n-1)/2$ pairs of bits at different positions in two different individuals; and $m(m-1)n/2$ pairs of bits located at the same position in two distinct individuals. Hence the result, by the same reasoning as in theorem 1.3. ∎

ACKNOWLEDGMENTS

I would like to thank Pierre Pansu for introducing me to the subject of genetic algorithms, for weekly conversations and numerous comments. Many thanks to Raphaël Cerf for having organised a small workshop on genetic algorithms I attended, and for helpful conversations. Thanks to Claire Kenyon as well for her presenting the work of Y. Rabani, Y. Rabinovich and A. Sinclair and for detailed comments on the manuscript. Thanks to the referee for useful comments.

REFERENCES

[1] S. Arora, Y. Rabani, U. Vazirani, *Simulating quadratic dynamical systems is PSpace-complete*, Proc 26th ACM Symp Theory of Computing (1994), p. 459–467.

[2] W. J. Ewens, *Mathematical Population Genetics*, Springer-Verlag, Berlin (1979).

[3] R.A. Fisher, *The genetical theory of natural selection*, Clarendon Press, Oxford (1930).

[4] S. Karlin, H.Taylor, *A second course in stochastic processes*, Academic Press, New York (1981).

[5] Y. Rabani, Y. Rabinovich and A. Sinclair, *A computational view of population genetics*, Random Structures and Algorithms **12** (1998), 4, p. 313–334.

[6] Y. Rabinovich, A. Sinclair, A. Widgerson, *Quadratic dynamical systems*, Proc 23rd IEEE Symp Foundations of Computer Science (1992), p. 304–313.

[7] S. Wright, *Evolution in Mendelian populations*, Genetics **16** (1931), p. 97–159.