The Annals of Applied Probability
2003, Vol. 13, No. 4, 1534–1568

SHARP ASYMPTOTIC RESULTS FOR SIMPLIFIED MUTATION–SELECTION ALGORITHMS

BY J. BÉRARD AND A. BIENVENÜE

Université Lyon I and LMC–ENSIMAG

We study the asymptotic behavior of a mutation–selection genetic algorithm on the integers with finite population of size $p \geq 1$. The mutation is defined by the steps of a simple random walk and the fitness function is linear. We prove that the normalized population satisfies an invariance principle, that a large-deviations principle holds and that the relative positions converge in law. After n steps, the population is asymptotically around $\sqrt{n}$ times the position at time 1 of a Bessel process of dimension $2p - 1$.

1. Introduction.

1.1. *Motivation.* Holland [9] introduced genetic algorithms as an optimization method, inspired by a biological analogy with (what one assumes to be) the mechanisms of Darwinian evolution. They are now a popular tool for solving hard combinatorial optimization problems. In these algorithms, a finite population of particles evolves under the action of three operators:

- Selection randomly resamples the population. Particles with higher fitness are more likely to be selected, and there is a tendency to eliminate the particles with lower fitness.
- Mutation randomly modifies each particle.
- Mating creates a new population of "offspring" from pairs of particles of the previous population.

In a combinatorial optimization context, particles are feasible solutions, and the fitness of a particle is the function to be maximized. Thus, the selection operator directs the evolution toward an increase in the fitness, while mutation and mating preserve the population diversity and allow the algorithm to visit large parts of the space of solutions.

Despite the successes that genetic algorithms encounter in practical applications and the numerous experimental studies of their properties, there exist few rigorous results on their behavior. Among the exceptions is Cerf [5], who obtained asymptotic convergence results for genetic algorithms with rare transitions. See also Rabinovich and Wigderson [13], who studied the convergence speed of several

Received March 2001; revised September 2002.
AMS 2000 subject classifications. 60F05, 60F10, 60F17, 92D15.
Key words and phrases. Genetic algorithm, invariance principle, large-deviations, population dynamics, random walks, interacting particle systems.

genetic algorithms defined on binary strings and the rate of escape of algorithms defined on the integers.

The algorithms we study belong to this latter category. We deliberately choose a very simplified context to get a detailed understanding of the effects of selection through a fully rigorous mathematical treatment. We take the fitness function to be $f(x) = x$, and our algorithm does not use a mating operator, but mutation and selection only. The price to pay is that our results do not apply to complex optimization situations. However, this simple model has been used in biology to study the evolution of a population of viruses; see [10] and [15]. Moreover, Bonnaz [4] applied theoretical physics methods to get predictions about its behavior (asymptotics of the mean and of the variance of the population) and supported them by numerical simulations. Our results confirm these predictions.

1.2. *Description of the model and notation.* A mutation–selection genetic algorithm on the integers with finite population of size $p \geq 1$ is specified by a simple random walk and by the fitness function $f(x) = x$, as follows.

Let $\mathbb{N}$ denote the set of nonnegative integers $\{0, 1, 2, \dots\}$. The algorithm is defined as a Markov chain $X_n = (X_n^{(i)})_{1 \leq i \leq p}$ with state space $\mathbb{N}^p$ and starting state $X_0 = (1, \dots, 1)$. The particles (or individuals) $X_n^{(1)}, \dots, X_n^{(p)}$ comprise the population at time n. Thus, the number of particles is constant equal to p and does not vary during the time evolution of the population. Note that $X_n^{(i)}$ stands for both the location of the individual and its fitness, since we consider the case $f(x) = x$.

The transitions of the Markov chain X_n are defined as follows:

1. *Selection step*: $X_n \to Y_n$. If $X_n = (0, \dots, 0)$, then $Y_n = (1, \dots, 1)$. Else, each $Y_n^{(i)}$, $1 \leq i \leq p$, is chosen randomly and independently of the others in the set $\{X_n^{(i)}, 1 \leq i \leq p\}$ according to the probability law
$$\frac{1}{S_n} \sum_{i=1}^{p} X_n^{(i)} \delta_{X_n^{(i)}} \qquad \text{where } S_n = \sum_{i=1}^{p} X_n^{(i)}.$$
Note that this procedure ensures that all $Y_n^{(i)}$'s are different from 0. Thus, the choice of $Y_n^{(i)}$ among the $X_n^{(i)}$'s is biased in favor of those $X_n^{(i)}$ that have the larger values. The weight of $X_n^{(i)}$ in this resampling process is proportional to its value, so we shall refer to this selection procedure as *proportional selection*, as opposed to the *uniform* or *neutral selection* procedure, which would consist of resampling the $X_n^{(i)}$'s with equal weights $1/p$, regardless of their actual values.
2. *Mutation step*: $Y_n \to X_{n+1}$. Each particle $Y_n^{(i)}$ evolves independently of the others and performs one step of a simple random walk on $\mathbb{N}$ (symmetric, to the nearest neighbors). The new positions are $X_{n+1}^{(i)}$, $1 \leq i \leq p$.

Note that, at a given time, several particles may assume the same value. The choice $X_n = (0, \dots, 0) \to Y_n = (1, \dots, 1)$ is arbitrary. The point here is to prevent particles from taking negative values. Other choices for the mutation process are possible.

We shall also consider the *neutral selection chain*, defined as a Markov chain $V_n = (V_n^{(i)})_{1\le i\le p}$ with state space $\mathbb{N}^p$ and starting state $V_0 = (0, \dots, 0)$, by the following transitions.

1. *Neutral selection step*: $V_n \to W_n$. Each $V_n^{(i)}$, $1 \le i \le p$, is chosen randomly and independently of the others in the set $\{V_n^{(i)}, 1 \le i \le p\}$ according to the probability law
$$\sum_{i=1}^{p} \frac{1}{p} \delta_{V_n^{(i)}}.$$
2. *Mutation step*: $W_n \to V_{n+1}$. Each particle $W_n^{(i)}$ evolves independently of the others and performs one step of a simple symmetric random walk on $\mathbb{N}$ with boundary transition probabilities $p_{0\to 0} = p_{0\to 1} = 1/2$. The new positions are $V_{n+1}^{(i)}$, $1 \le i \le p$.

We denote by $\mathcal{F}_n$ and $\mathcal{G}_n$ two σ-algebras:
$$\mathcal{F}_n = \sigma(X_0, Y_0, \dots, X_{n-1}, Y_{n-1}, X_n),$$
and $\mathcal{G}_n$ is the σ-algebra generated by $\mathcal{F}_n \cup \sigma(Y_n)$. We set
$$S_n = \sum_{i=1}^{p} X_n^{(i)}, \qquad T_n = \sum_{i=1}^{p} Y_n^{(i)},$$
$$F_n = S_n/p, \qquad G_n = T_n/p, \qquad \Delta F_n = F_{n+1} - F_n.$$
If $x \in \mathbb{N}^p$, E^x denotes the expectation under the law of the Markov chain X starting from $X_0 = x$. The barycenter $m(x)$ and the diameter $d(x)$ of x are
$$m(x) = (x_1 + \cdots + x_p)/p, \qquad d(x) = \sup_{1\le i,j\le p} |x_i - x_j|.$$
Finally, $M_n(t) = X_{\lfloor nt\rfloor}/\sqrt{n}$ and $Z_n(t) = m(M_n(t)) = F_{\lfloor nt\rfloor}/\sqrt{n}$.

1.3. *Statement of the results.* For any fixed p, we describe the asymptotic behavior of (X_n), when n goes to ∞, as follows.

THEOREM 1. *The sequence of processes $(M_n)_{n\ge 0}$ converges in law in the Skorohod space $\mathcal{D}([0,1], \mathbb{R}^p)$ to $(1, \dots, 1)R$, where R is a Bessel process of dimension $(2p-1)$, starting from* 0.

THEOREM 2. *For all i, the sequence $(X_n^{(i)}/n)$ satisfies a large-deviations principle on $[0,1]$ with good rate function I, where I is the rate function associated with a simple reflected random walk on $\mathbb{N}$, that is,*
$$I(a) = \tfrac{1}{2}[(1-a)\log(1-a) + (1+a)\log(1+a)].$$

To state our last result, say that two elements $x, y \in \mathbb{N}^p$ are equivalent whenever $x - y$ is an integer multiple of $(1, \dots, 1)$ and denote by $\Theta(x)$ the class of x for this equivalence relation. Thus, $\Theta(X_n)$ informs us of the *relative* positions of the particles in the cloud X_n. Theorem 3 essentially says that these relative positions converge in law and that performing neutral selection instead of proportional selection would lead to the same limiting law.

THEOREM 3. *$\Theta(X_n)$ converges in law to a law π_p, which is also the limiting law in the case of neutral selection.*

(See the remarks about the description of π_p in Section 6.)

Theorem 1 has been announced in [2], and weaker versions of Theorems 1 and 2 have been announced in [1].

First, note that, since there is always a positive probability bounded away from 0 of choosing p times the same value in the selection step, the diameter roughly does not grow with n. This explains why the whole population X_n behaves like a single point on the $\sqrt{n}$ scale. Note that the $\sqrt{n}$ growth rate of the fitness is the same with proportional or neutral selection for any population size p. Neutral selection would lead, in the limit, to reflected Brownian motion. However, as p increases, the limiting law for X_n is more and more different from the reflected normal law that would hold with neutral selection. It gives more and more weight to the large values. Moreover, on the n scale, the large-deviations principle is identical to the case of a simple reflected random walk. Thus, Theorems 2 and 3 imply that in the limit, as $n \to +\infty$, selection has no effect on X_n/n or $\Theta(X_n)$.

The behavior of the infinite-population version of this model is strikingly different; see [12]. In the infinite-population model, the fitness grows linearly, as opposed to our $\sqrt{n}$ scaling factor. This is due to the fact that, in the infinite-population version, the probability of choosing the same value for all the components is 0, whereas in the finite-population version, this probability is bounded away from 0 and essentially forces the process to become closer and closer to neutral selection.

1.4. *Contents.* In Section 2, we prove that the cloud of particles goes to ∞ in probability, and we give a stochastic upper bound of the diameter of the cloud. Section 3 collects technical lemmas. Sections 4 and 5 are devoted to the detailed exposition of the invariance principle for the normalized cloud (Theorem 1) and of the large-deviations principle (Theorem 2). In Section 6, we prove the convergence in law of the nonnormalized cloud of particles (Theorem 3). In Section 7, we present numerical simulations of the algorithm.

2. Two preliminary results. From Lemma 1, the cloud of particles X_n is stochastically bounded from below by a neutral selection chain V_n. For fixed n and i, $V_n^{(i)}$ follows the same law as the position at time n of a simple reflected random walk. However, for fixed i, the component process $(V_n^{(i)})_n = (V_0^{(i)}, V_1^{(i)}, V_2^{(i)}, \dots)$ itself is *not* a simple random walk. Lemma 2 provides a stochastic upper bound on the diameter of X_n.

LEMMA 1. *Let $x \in \mathbb{N}^p$. It is possible to define on the same probability space two p-particle processes $X_n = (X_n^{(1)}, \dots, X_n^{(p)}) \in \mathbb{N}^p$ and $V_n = (V_n^{(1)}, \dots, V_n^{(p)}) \in \mathbb{N}^p$, such that*:

- *the process (X_n) follows the law of the algorithm under study*, *started at x*, *as defined in Section* 1.2;
- *the process (V_n) follows the law of the neutral selection chain*, *started at* 0, *as defined in Section* 1.2;
- $V_n^{(i)} \le X_n^{(i)}$, *almost surely.*

As a consequence, for fixed n and i, $V_n^{(i)}$ follows the law of the position at time n of a simple symmetric random walk on $\mathbb{N}$ starting from 0 with transition probabilities $p_{0\to 0} = p_{0\to 1} = 1/2$. By the null recurrence of such a random walk, we get the following result.

COROLLARY 1. *For fixed $1 \le i \le p$, $X_n^{(i)}$ goes to ∞ in probability*, *that is*,

$$\forall K > 0, \qquad \mathrm{P}\big[X_n^{(i)} \le K\big] \underset{n\to+\infty}{\longrightarrow} 0.$$

PROOF OF LEMMA 1. We design a coupling between the process $(X_n)_{n\ge 0}$ and another mutation–selection process $(V_n)_{n\ge 0}$ with the same mutation steps but with *neutral* selection, started at 0. It should be intuitively clear that (X_n) dominates (V_n) in some sense, since the selection steps of (X_n) favor individuals with large values, whereas the selection steps of (V_n) give equal weight to all individuals, regardless of their respective values.

Before giving precise definitions, let us give an informal description of how our coupling argument works. We start with two p-tuples X_n and V_n such that, for all $1 \le i \le p$, $V_n^{(i)} \le X_n^{(i)}$, and we want to prove that we may simultaneously apply a proportional selection step to X_n, $X_n \to Y_n$, and a neutral selection step to V_n, $V_n \to W_n$, so as to preserve the ordering and get $W_n^{(i)} \le Y_n^{(i)}$ for all $1 \le i \le p$. For the sake of simplicity, assume that all the individuals $X_n^{(i)}$ have distinct values.

Fix an index $1 \le i \le p$. It is not hard to check that, for all $1 \le k \le p$, the probability that $Y_n^{(i)}$ is chosen among the k individuals in X_n that have the smallest values is less than $k \times 1/p$. Note that $k \times 1/p$ is the probability that, during a neutral selection step, $W_n^{(i)}$ is chosen among the $V_n^{(j)}$'s whose indices j are precisely those of the k smallest individuals in X_n. Thus, it may be clear that we can couple the proportional selection step $X_n \to Y_n$ and the neutral selection step $V_n \to W_n$ in such a way that, whenever $Y_n^{(i)}$ is chosen to be, say, the kth smallest individual in X_n, $W_n^{(i)}$ is chosen to be one of the $V_n^{(j)}$'s whose indices j are precisely those of the k smallest individuals in X_n. Since all those $V_n^{(j)}$'s have a smaller value than the kth smallest individual in X_n, owing to the fact that $V_n^{(g)} \le X_n^{(g)}$ for all g, such a coupling preserves the ordering and leads to *p-tuples* Y_n and W_n such that, for all i, $W_n^{(i)} \le Y_n^{(i)}$.

Let us now give a formal description. We start with $V_0 = (0, \dots, 0)$ and $X_0 = x \in \mathbb{N}^p$, so the ordering condition $V_0^{(i)} \le X_0^{(i)}$ for all $1 \le i \le p$ is obviously met. Assume that the coupling has already been constructed with the required properties up to time n.

We explicitly express the selection steps $X_n \to Y_n$ and $V_n \to W_n$ with the help of p uniform random variables $U_1, \dots, U_p$ independent and independent from the past. Denote by $x_1 = X_n^{(q_1)} \le \cdots \le x_p = X_n^{(q_p)}$ the ordered values of the p-tuple X_n and let $s = \sum_{i=1}^p x_i$. When $X_n \ne (0, \dots, 0)$, the individuals $Y_n^{(i)}$ are chosen as

$$Y_n^{(i)} = x_{k_i},$$

where, for all $1 \le i \le p$, k_i is uniquely defined by

$$\sum_{j=1}^{k_i - 1} x_j < sU_i \le \sum_{j=1}^{k_i} x_j \qquad \forall i = 1, \dots, p.$$

When $X_n = (0, \dots, 0)$, we set $Y_n = (1, \dots, 1)$.

We easily check that this construction is consistent with the definition of the process (X_n):

- Conditional on X_n, the $Y_n^{(i)}$ are independent, since the definition of k_i refers only to X_n and the random variable U_i.
- Conditional on X_n, for all $1 \le r \le p$, the probability that $Y_n^{(i)}$ is chosen as $X_n^{(q_r)}$ equals

$$s^{-1} \sum_{j=1}^{r} x_j - s^{-1} \sum_{j=1}^{r-1} x_j = s^{-1} x_r = \frac{1}{S_n} X^{(q_r)}.$$

Let us now express the neutral selection step $V_n \to W_n$ in terms of the *same* uniform random variables U_i.

We reindex the $V_n^{(i)}$ according to the *same* $q_1, \dots, q_p$ that were defined by the reordering of the $X_n^{(i)}$: $v_1 = V_n^{(q_1)}, \dots, v_p = V_n^{(q_p)}$.

The individual $W_n^{(i)}$ is chosen as

$$W_n^{(i)} = v_{l_i},$$

where l_i is uniquely defined by

$$l_i - 1 < pU_i \le l_i \qquad \forall\, i = 1, \dots, p.$$

Again, we check that this construction is consistent with the definition of the process (V_n):

- Conditional on V_n and X_n, the $W_n^{(i)}$ are independent since the definition of l_i refers only to the random variable U_i.
- Conditional on V_n and X_n, for all $1 \le r \le p$, the probability that $W_n^{(i)}$ is chosen as $V_n^{(q_r)}$ equals $1/p$.

Conditioning with respect to X_n is a priori necessary in the above argument, since the indices $q_1, \dots, q_p$ depend on X_n and were used in the definition of the $W_n^{(i)}$. However, the above argument shows that, conditional on V_n and X_n, all choices $W_n = (V_n^{(i_1)}, \dots, V_n^{(i_p)})$ are equally likely, with common probability $1/p^p$. Thus, the conditional law of W_n with respect to both V_n and X_n is, in fact, independent from X_n, and our construction is consistent with the definition of V_n.

Let us now prove that the ordering is preserved by our definition of the selection steps. Assume first that $X_n \neq (0, \dots, 0)$. The key observation is that, since $(x_i)_i$ is a nondecreasing sequence, l_i must be less than k_i for all $1 \le i \le p$. As a consequence, $x_{l_i} \le x_{k_i}$. Since we assume that, for all i, $V_n^{(i)} \le X_n^{(i)}$, v_l is less than x_l for all l. Putting these inequalities together, we get

$$W_n^{(i)} = v_{l_i} \le x_{l_i} \le x_{k_i} = Y_n^{(i)}.$$

When $X_n = (0, \dots, 0)$, V_n must also be equal to $(0, \dots, 0)$ since we assume that $V_n^{(i)} \le X_n^{(i)}$ for all i. In this case, $Y_n = (1, \dots, 1)$ by definition, and $W_n = (0, \dots, 0)$, so we again have that $W_n^{(i)} \le Y_n^{(i)}$ for all i.

Let us now define the mutation steps.

Let $\epsilon_1, \dots, \epsilon_p$ be independent random variables, independent from the past of symmetric ± 1 Bernoulli law:

$$\begin{aligned} X_{n+1}^{(i)} &= \left(Y_n^{(i)} + \epsilon_i\right), \\ V_{n+1}^{(i)} &= \sup\left\{W_n^{(i)} + \epsilon_i, 0\right\}. \end{aligned}$$

It is routine to check that this construction is consistent with the definitions of X_n and V_n and that $V_{n+1}^{(i)} \le X_{n+1}^{(i)}$.

We have thus designed a step-by-step coupling between (X_n) and (V_n) such that the ordering is preserved. $\square$

The proof of the next lemma is based on the following crucial tool.

DEFINITION 1. For $n \geq 0$, call n a crunch time if the population Y_n is concentrated on one point. That is, the selection step $X_n \to Y_n$ amounts to choosing p times the same value.

LEMMA 2. *Let $X_0 = x \in \mathbb{N}^p$. For all $n \in \mathbb{N}^*$, there is a random variable A_n of geometric law on $\mathbb{N}^*$ with parameter p^{1-p}, such that*

$$d(X_n) \leq 2A_n \mathbf{1}_{\{A_n \leq n\}} + \big(d(x) + 2n\big)\mathbf{1}_{\{A_n > n\}}.$$

As a special case, if $d(x) = 0$,

$$d(X_n) \leq 2A_n.$$

PROOF. The idea of the proof is the following: the probability that n is a crunch time is strictly bounded away from 0. Hence, looking backwards in time, the random time interval between n and the last crunch time before n has subgeometric tail. Since a mutation step does not increase the diameter by more than a constant, a bound on the diameter at time n follows.

Given $\mathcal{F}_n$, the conditional probability q_n that n is a crunch time satisfies

$$q_n \geq \frac{1}{S_n^p} \sum_{k=1}^{p} (X_n^{(k)})^p.$$

Indeed, if $X_n = (0, \ldots, 0)$, then $q_n = 1$. Else, the probability of selecting p times the index j is $(X_n^{(j)})^p / S_n^p$. In both cases, $q_n \geq p^{1-p}$. Thus, one may assume that, conditional on X_n, the proportional selection step $X_n \to Y_n$ is built owing to random variable U_n with uniform law on [0, 1], such that:

- for all $1 \leq k \leq p$, if
$$\sum_{i=1}^{k-1} (X_n^{(i)}/S_n)^p < U_n \leq \sum_{i=1}^{k} (X_n^{(i)}/S_n)^p,$$
then $Y_n^{(1)} = \cdots = Y_n^{(p)} = X_n^{(k)}$;
- if
$$\sum_{i=1}^{p} (X_n^{(i)}/S_n)^p < U_n \leq 1,$$
then the $Y_n^{(i)}$'s are chosen among the $X_n^{(i)}$'s consistently with the definition of proportional selection (an explicit definition is not needed).

Consider now the algorithm up to time n and assume that the proportional selection steps are built using i.i.d. uniform random variables $U_0, \dots, U_n$ as explained above. Define B_n by

$$B_n = n - \sup\big(\{k < n,\ U_k \le p^{1-p}\} \cup \{-1\}\big).$$

The above construction yields

$$\{B_n \le n\} \subset \{d(Y_{n-B_n}) = 0\}.$$

Moreover, the independence of the random variables U_n implies that B_n may be written as $B_n = \min\{A_n, (n+1)\}$, where A_n follows a geometric law on $\mathbb{N}^*$ with parameter p^{1-p}. A selection step does not increase the diameter of the cloud of particles. A mutation step does not increase the diameter by more than 2. Hence,

- if $B_n \le n$ (i.e., if $A_n \le n$), then $d(X_n) \le 2B_n$;
- if $B_n = n+1$ (i.e., if $A_n > n$), then $d(X_n) \le d(x) + 2n$. $\square$

COROLLARY 2. *For all $x \in \mathbb{N}^p$ and $k \in \mathbb{N}$, there is a constant $\alpha_{p,k}$ such that, for all $n \ge 0$, the algorithm started at $X_0 = x$ satisfies*

$$\mathrm{E}^x d(X_n)^k \le d(x)^k + \alpha_{p,k}.$$

3. Technical lemmas. Lemma 3 computes $\mathrm{E}^x(\Delta F_n|\mathcal{F}_n)$ and $\mathrm{E}^x((\Delta F_n)^2|\mathcal{F}_n)$.

LEMMA 3. *For all $x \in \mathbb{N}^p$,*

$$\mathrm{E}^x(\Delta F_n|\mathcal{F}_n) = \mathbf{1}_{\{X_n \ne 0\}} \frac{1}{p^2 F_n} \sum_{1 \le i < j \le p} \big(X_n^{(j)} - X_n^{(i)}\big)^2 + \mathbf{1}_{\{X_n = 0\}}$$

and

$$\begin{aligned}\mathrm{E}^x\big((\Delta F_n)^2\big|\mathcal{F}_n\big) = p^{-1} &+ p^{-2} \sum_{1 \le i \ne j \le p} \big[\mathbf{1}_{\{X_n \ne 0\}} H_{i,j}(X_n) + \mathbf{1}_{\{X_n = 0\}}\big] \\ &+ p^{-2} \sum_{1 \le i \le p} \big[\mathbf{1}_{\{X_n \ne 0\}} K_i(X_n) + \mathbf{1}_{\{X_n = 0\}}\big],\end{aligned}$$

where we used the following shorthand:

$$H_{i,j}(X_n) = \sum_{1 \le a,b \le p} \frac{X_n^{(a)} X_n^{(b)}}{S_n^2} \big(X_n^{(a)} - X_n^{(i)}\big)\big(X_n^{(b)} - X_n^{(j)}\big),$$

$$K_i(X_n) = \sum_{1 \le a \le p} \frac{X_n^{(a)}}{S_n} \big(X_n^{(a)} - X_n^{(i)}\big)^2.$$

PROOF. We prove the formula for $(\Delta F_n)^2$ and omit the easier proof of the formula for ΔF_n. Note that

$$\mathrm{E}^x\big((\Delta F_n)^2|\mathcal{G}_n\big) = \mathrm{E}\big((G_n - F_n + (\epsilon_1 + \cdots + \epsilon_p)/p)^2|\mathcal{G}_n\big),$$

where $(\epsilon_i)_i$ is i.i.d., of symmetric ± 1 Bernoulli law and independent of $\mathcal{G}_n$. Hence,

$$\mathrm{E}^x\big((\Delta F_n)^2|\mathcal{G}_n\big) = (G_n - F_n)^2 + p^{-1}.$$

Conditioning again by $\mathcal{F}_n$, we get the result of the lemma by noting that

$$\mathrm{E}^x\big((G_n - F_n)^2|\mathcal{F}_n\big) = p^{-2} \sum_{1 \le i,j \le p} \mathrm{E}^x\big((Y_n^{(i)} - X_n^{(i)})(Y_n^{(j)} - X_n^{(j)})|\mathcal{F}_n\big). \qquad \square$$

Lemma 4 states that the transition probabilities of the selection step can be approximated by those corresponding to a neutral selection. Lemma 2 means roughly that the diameter $d(X_n)$ remains uniformly bounded. Since, by Corollary 1, S_n goes to ∞, this means that the difference between any two transition probabilities $X_n^{(i)}/S_n$ and $X_n^{(j)}/S_n$ is small on average.

LEMMA 4. *Let $x = (x(1), \dots, x(p))$ be any p-tuple of nonnegative real numbers, with $x \neq 0$. Set $S = x(1) + \cdots + x(p)$. For k and $i_1, \dots, i_k$ in $\{1, \dots, p\}$ and for $n \ge 0$,*

$$\left|\left(\prod_{j=1}^{k} \frac{x(i_j)}{S}\right) - \frac{1}{p^k}\right| \le \frac{kd(x)}{S}.$$

PROOF. By induction on k. If $k = 1$,

$$\left|\frac{x(i)}{S} - \frac{1}{p}\right| = \left|\frac{\sum_{j=1}^{p}(x(i) - x(j))}{pS}\right| \le \frac{d(x)}{S}.$$

If the result holds for all $l \le k-1$ with $k \ge 2$, one has, reducing to the same denominator,

$$\left|\prod_{l=1}^{k} \frac{x(i_l)}{S} - \frac{1}{p^k}\right| = \frac{1}{(pS)^k}\left|\sum_{1 \le j_1, \dots, j_k \le p} \left(\prod_{l=1}^{k} x(i_l) - \prod_{l=1}^{k} x(j_l)\right)\right|.$$

Adding and subtracting $x(i_1) \times \cdots \times x(i_{k-1}) \times x(j_k)$ to each term of this sum, we get

$$\left|\prod_{l=1}^{k} \frac{x(i_l)}{S} - \frac{1}{p^k}\right| = \frac{1}{(pS)^k}\left|\sum_{1 \le j_1, \dots, j_k \le p} \left[\big(x(i_k) - x(j_k)\big)\prod_{l=1}^{k-1} x(i_l) + x(j_k)\left(\prod_{l=1}^{k-1} x(i_l) - \prod_{l=1}^{k-1} x(j_l)\right)\right]\right|,$$

which leads to the bound

$$\begin{aligned}\left|\prod_{l=1}^{k}\frac{x(i_l)}{S}-\frac{1}{p^k}\right| &\le \left(\prod_{l=1}^{k-1}\frac{x(i_l)}{S}\right)\left|\sum_{j_k=1}^{p}\frac{x(i_k)-x(j_k)}{pS}\right| \\ &\qquad + \frac{x(j_k)}{S}\left|\sum_{1\le j_1,\dots,j_{k-1}\le p}\left(\prod_{l=1}^{k-1}x(i_l)-\prod_{l=1}^{k-1}x(j_l)\right)\right| \\ &\le \left|\frac{x(i_k)}{S}-\frac{1}{p}\right|+\left|\left(\prod_{l=1}^{k-1}\frac{x(i_l)}{S}\right)-\frac{1}{p^{k-1}}\right| \\ &\le \frac{(1+(k-1))d(x)}{S}\end{aligned}$$

by the induction hypothesis. □

We omit the proof of Lemma 5.

LEMMA 5. *Let* $\alpha \in]-1,1[$ *and* $q \in \mathbb{N}$. *As* r *goes to* ∞,

$$\sum_{k=0}^{r-1}\alpha^{r-1-k}k^q = \frac{1}{1-\alpha}r^q + o(r^q).$$

4. Invariance principle. The following proposition is the key result for our proof of the invariance principle. It gives asymptotic estimates of the even moments of F_n that hold uniformly with respect to the initial value $X_0 = x$ of the chain. Theorem 1 follows, using the Markov property of X and the method of moments. We use the following definition in the statement of the proposition.

DEFINITION 2. For $h \in \mathbb{N}$, the function $(x,n)\mapsto \beta(x,n,h)$, defined on $n \in \mathbb{N}$ and $x \in \mathbb{N}^p$, satisfies the property (Q_h) if $\beta(x,n,h)$ is the sum of a finite number of terms of the form

$$\epsilon(x,n)m(x)^i n^j d(x)^k \qquad \text{with } i,j,k\in\mathbb{N},\ 2j+i\le 2h,$$

where $\epsilon(x,n)$ is bounded and goes to 0 as n goes to ∞ uniformly with respect to x.

PROPOSITION 1. *For* $h \in \mathbb{N}$, *the following property holds, which we denote by* (P_h). *For* $x \in \mathbb{N}^p$,

$$\mathrm{E}^x F_n^{2h} = f(x,n,h) + \alpha(x,n,h),$$

where $\alpha(x,n,h)$ *satisfies* (Q_h) *and*

$$f(x,n,h) = \sum_{i=0}^{h} f_i(h)m(x)^{2i}n^{h-i} \qquad \textit{with } f_i(h)\in\mathbb{R}.$$

The numbers $f_i(h)$ *are uniquely determined by the following induction equations*:

$$\begin{aligned} &f_h(h) = 1, \\ (4.1) \qquad &f_i(h) = \frac{2h(p-1) + h(2h-1)}{h-i} f_i(h-1), \qquad 0 \le i \le h-1. \end{aligned}$$

Our proof of Proposition 1 is very technical. We defer it to Appendix A and sketch the idea now. We use induction on q. The main step is to prove that, as $n \to +\infty$, the difference

$$\mathrm{E}F_{n+1}^{2q} - \mathrm{E}F_n^{2q} = \mathrm{E}(F_n + \Delta F_n)^{2q} - \mathrm{E}F_n^{2q} \tag{4.2}$$

involves only the expectations of the random variables F_r^{2h} with $r \le n$ and $h \le q-1$. To see this, we show that, expanding the $2q$th power in (4.2), the leading terms when $n \to +\infty$ are those involving ΔF_n and $(\Delta F_n)^2$ only. Using the approximation of the transition probabilities of the selection step by uniform ones (see Lemma 4), we thus express (4.2) in terms of expectations of F_n^{2h} and of

$$\varphi_h(X_n) = \big(X_n^{(a)} - X_n^{(b)}\big)^2 F_n^{2h}$$

for $h \le q-1$. The asymptotic behavior of quantities of the last type stems from (P_h) and from an approximate contraction property of the selection operator which reads as follows:

$$\mathrm{E}\varphi_h(X_{n+1}) = \lambda \mathrm{E}\varphi_h(X_n) + \psi_h\big[\mathrm{E}(F_r^{2i}); r \le n, i \le q-1\big]\big(1 + o(1)\big),$$

where $0 < \lambda < 1$ and ψ_h is explicit.

REMARK 1. The details of the proof of Proposition 1 yield that $\alpha(x, n, h)$ contains no term of the form $\epsilon(x, n) m(x)^{2h} d(x)^k$.

The numbers $f_i(h)$ in Proposition 4.1 are related to the $(2p-1)$-dimensional Bessel process $(R(t))_{0 \le t \le 1}$ starting from 0. The law of $R(t)$ is as follows.

LEMMA 6. *For* $u \in \mathbb{R}_+$,

$$\mathrm{E}^u R(t)^{2h} = \sum_{i=0}^{h} f_i(h) u^{2i} t^{h-i}.$$

PROOF. Let $d = 2p - 1$ and $\nu = d/2 - 1 = p - 3/2$. Using the explicit expression of the density of the Bessel semigroup density (see [14], page 415) and formulas 6.631 and 9.212 of [8], we get

$$\begin{aligned}
\mathrm{E}^u R(t)^{2h} &= \int_0^\infty t^{-1}\left(\frac{y}{u}\right)^\nu y^{2h+1} \exp\left(-\frac{u^2+y^2}{2t}\right) I_\nu\left(\frac{uy}{t}\right) dy \\
&= t^{-1} u^{-\nu} \exp\left(-\frac{u^2}{2t}\right) \frac{(u/t)^\nu \Gamma(h+\nu+1)}{2^{\nu+1}(2t)^{-(h+\nu+1)}\Gamma(\nu+1)} \\
&\quad \times \Phi\left(h+\nu+1, \nu+1; \frac{u^2}{2t}\right) \\
&= (2t)^h \frac{\Gamma(h+\nu+1)}{\Gamma(\nu+1)} \Phi\left(-h, \nu+1; -\frac{u^2}{2t}\right) \\
&= \sum_{i=0}^{h} g_i(h) u^{2i} t^{h-i},
\end{aligned}$$

where $\Phi = {}_1F_1$ is the Kummer confluent hypergeometric function and I_ν is the modified Bessel function of the first kind (see [8] for precise definitions). By the definition of ${}_1F_1$, one has

$$g_0(h) = d \times (d+2) \times \cdots \times \bigl(d + 2(h-1)\bigr)$$

and, for $1 \le i \le h$,

$$\begin{aligned}
g_i(h) &= g_0(h) \frac{h \times (h-1) \times \cdots \times (h-i+1)}{(\nu+1) \times (\nu+2) \times \cdots \times (\nu+1+i-1)} \frac{1}{2^i i!} \\
&= g_0(h) \frac{h \times (h-1) \times \cdots \times (h-i+1)}{d \times (d+2) \times \cdots \times (d+2(i-1))} (i!)^{-1} \\
&= \binom{h}{i} \prod_{i \le j \le h-1} (d+2j).
\end{aligned}$$

Since the $g_i(h)$ satisfy (4.1), we are done. □

The next two propositions study the finite-dimensional marginals of Z_n.

PROPOSITION 2. *The finite-dimensional distributions of the process Z_n converge to those of R.*

PROOF. We prove by induction on m that, for $t = (t_1, \dots, t_m)$, where the t_i's are nonnegative real numbers with $t_i < t_{i+1}$, and for $h = (h_1, \dots, h_m)$, h_i positive integers, as n goes to ∞,

$$\mathrm{E}\bigl(Z_n(t_1)^{2h_1} \times \cdots \times Z_n(t_m)^{2h_m}\bigr) \to \mathrm{E}\bigl(R(t_1)^{2h_1} \times \cdots \times R(t_m)^{2h_m}\bigr).$$

For $m = 1$, $\mathrm{E}Z_n(t)^h = n^{-h}\mathrm{E}F_{\lfloor nt \rfloor}^{2h}$, that is,

$$\mathrm{E}Z_n(t)^h = n^{-h}(f+\alpha)(0, \lfloor nt \rfloor, h) = t^h f_0(h) + o(1),$$

according to Proposition 1 and Corollary 2. According to Lemma 6, $\mathrm{E}Z_n(t)^h$ converges to $\mathrm{E}R(t)^h$ as n goes to ∞. Assume now that the result holds for $i \le m$, where $m \ge 1$. For $x \in \mathbb{N}^p$, r and $i \in \mathbb{N}$, set $\phi(x, r, i) = \mathrm{E}^x(F_r^{2i})$. Fix $h = (h_1, \dots, h_{m+1})$ and set

$$F(n, m) = F_{\lfloor nt_1 \rfloor}^{2h_1} \times \cdots \times F_{\lfloor nt_m \rfloor}^{2h_m} n^{-(h_1+\cdots+h_m)}.$$

Then

$$\begin{aligned}
&\mathrm{E}\big(Z_n(t_1)^{2h_1} \times \cdots \times Z_n(t_{m+1})^{2h_{m+1}}\big)\\
&\quad = \mathrm{E}\big(F(n, m) \times F_{\lfloor nt_{m+1} \rfloor}^{2h_{m+1}}\big) n^{-h_{m+1}}\\
&\quad = \mathrm{E}\Big(F(n, m)\phi\big(X_{\lfloor nt_m \rfloor}, \lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor, h_{m+1}\big) n^{-h_{m+1}}\Big)\\
&\quad = \mathrm{E}\Big(F(n, m)(f + \alpha)\big(X_{\lfloor nt_m \rfloor}, \lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor, h_{m+1}\big) n^{-h_{m+1}}\Big)\\
&\quad = A_n + B_n,
\end{aligned}$$

say, where A_n and B_n are the contributions of f and α, respectively. According to Proposition 1, the error term B_n is a sum of terms of the form

$$\begin{aligned}
b_{n,i,j} = {} & \mathrm{E}F(n, m)\epsilon\big(X_{\lfloor nt_m \rfloor}, \lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor\big)\\
& \times F_{\lfloor nt_m \rfloor}^{i}(\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)^j \, d\big(X_{\lfloor nt_m \rfloor}\big)^k n^{-h_{m+1}},
\end{aligned}$$

where $\epsilon(u, r)$ goes to 0 as r goes to ∞ uniformly with respect to u, and $2j + i \le 2h_{m+1}$. According to Hölder's inequality,

$$\begin{aligned}
b_{n,i,j} \le {} & \epsilon(\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)\big(\mathrm{E}F(n, m)^2\big)^{1/2}\\
& \times (\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)^j n^{i/2 - h_{m+1}}\\
& \times \big(n^{-2i}\mathrm{E}F_{\lfloor nt_m \rfloor}^{4i}\big)^{1/4}\big(\mathrm{E}d\big(X_{\lfloor nt_m \rfloor}\big)^{4k}\big)^{1/4},
\end{aligned}$$

where $\epsilon(r)$ goes to 0 as r goes to ∞.

By the induction hypothesis, $\mathrm{E}F(n, m)^2$ and $n^{-2i}\mathrm{E}F_{\lfloor nt_m \rfloor}^{4i}$ converge. Moreover, according to Corollary 2, $\mathrm{E}d(X_{\lfloor nt_m \rfloor})^{4k}$ is bounded. Hence, each $b_{n,i,j}$ goes to 0 as n goes to ∞.

As regards the main term A_n,

$$\begin{aligned}
A_n &= \sum_{i=0}^{h_{m+1}} \mathrm{E}F(n, m) f_i(h_{m+1}) F_{\lfloor nt_m \rfloor}^{2i}(\lfloor nt_{m+1} \rfloor - \lfloor nt_m \rfloor)^{h_{m+1}-i} n^{-h_{m+1}}\\
&= \sum_{i=0}^{h_{m+1}} f_i(h_{m+1})\mathrm{E}F_{\lfloor nt_1 \rfloor}^{2h_1} \times \cdots \times F_{\lfloor nt_{m-1} \rfloor}^{2h_{m-1}} \times F_{\lfloor nt_m \rfloor}^{2(h_m+i)} n^{-(h_1+\cdots+h_m+i)}\\
&\qquad \times (t_{m+1} - t_m)^{h_{m+1}-i}\big(1 + o(1)\big).
\end{aligned}$$

Hence, according to the induction hypothesis, A_n converges to

$$\begin{aligned}
A_\infty &= \sum_{i=0}^{h_{m+1}} f_i(h_{m+1}) \mathrm{E} R(t_1)^{2h_1} \times \cdots \times R\big(t_{m-1}^{2h_{m-1}}\big) \\
&\qquad \times R(t_m)^{2(h_m+i)} (t_{m+1}-t_m)^{h_{m+1}-i} \\
&= \mathrm{E}\Bigg(R(t_1)^{2h_1} \times \cdots \times R\big(t_m^{2h_m}\big) \\
&\qquad \times \sum_{i=0}^{h_{m+1}} f_i(h_{m+1}) R(t_m)^{2i} (t_{m+1}-t_m)^{h_{m+1}-i} \Bigg).
\end{aligned}$$

According to Lemma 6,

$$\begin{aligned}
A_\infty &= \mathrm{E} R(t_1)^{2h_1} \times \cdots \times R\big(t_m^{2h_m}\big) \mathrm{E}^{R(t_m)} R(t_{m+1}-t_m)^{2h_{m+1}} \\
&= \mathrm{E} R(t_1)^{2h_1} \times \cdots \times R(t_{m+1})^{2h_{m+1}},
\end{aligned}$$

and we are done with the proof of the induction.

Now, the method of moments entails that, for all $m \geq 1$, the following convergence in law holds:

$$\big(\epsilon_1 F_n(t_1), \ldots, \epsilon_m F_n(t_m)\big) \xrightarrow{\mathcal{L}} \big(\epsilon_1 R(t_1), \ldots, \epsilon_m R(t_m)\big)$$

as n goes to ∞, and where the $\epsilon_i \in \{-1, +1\}$ are i.i.d. random variables, independent of F_n and R, following a symmetric ± 1 Bernoulli law. The limiting law is uniquely determined by its moments since, for all $0 \leq t \leq 1$ and $\lambda \in \mathbb{R}_+$, the moment-generating function $\mathrm{E}(e^{\lambda R_t})$ is finite valued. The random variables $F_n(t_i)$ and $R(t_i)$ being nonnegative, this implies finally that

$$\big(F_n(t_1), \ldots, F_n(t_m)\big) \xrightarrow{\mathcal{L}} \big(R(t_1), \ldots, R(t_m)\big). \qquad \square$$

PROPOSITION 3. *There exists a constant C_p such that, for all $0 \leq t_1 \leq t \leq t_2 \leq 1$,*

$$\mathrm{E}\big(Z_n(t) - Z_n(t_1)\big)^2 \big(Z_n(t) - Z_n(t_2)\big)^2 \leq C_p (t_2 - t_1)^{3/2}.$$

PROOF. For $x \in \mathbb{N}^p$ and $r, r_1, r_2 \in \mathbb{N}$, we define

$$\begin{aligned}
\psi_1(x, r) &= \mathrm{E}^x\big((F_r - F_0)^2\big), \\
\psi_2(x, r_1, r_2) &= \mathrm{E}^x\big((F_{r_1} - F_0)^2 \psi_1(X_{r_1}, r_2)\big).
\end{aligned}$$

According to the Markov property, for all integers $a < b < c$, we have

$$\mathrm{E}(F_b - F_a)^2 (F_b - F_c)^2 = \mathrm{E}\big(\psi_2(X_a, b-a, c-b)\big). \tag{4.3}$$

Since $(s_1 - s_2)^2 \le |s_1^2 - s_2^2|$ for nonnegative s_1 and s_2,

$$\psi_1(x,r) = \mathrm{E}^x (F_r - F_0)^2 \le \mathrm{E}^x |F_r^2 - F_0^2|.$$

The estimates of Proposition 1 and Remark 1 lead to the inequality:

$$\mathrm{E}^x |F_r^2 - F_0^2| \le c_p^{(1)} (1 + d(x)^2) r.$$

Hence,

$$\psi_2(x, r_1, r_2) \le \mathrm{E}^x \big((F_{r_1} - F_0)^2 c_p^{(1)} (1 + d(X_{r_1})^2) r_2 \big).$$

According to Hölder's inequality, we have

$$\psi_2(x, r_1, r_2) \le c_p^{(1)} r_2 \big(\mathrm{E}^x (F_{r_1} - F_0)^4\big)^{1/2} \big(\mathrm{E}^x (1 + d(X_{r_1})^2)^2\big)^{1/2}. \tag{4.4}$$

From Corollary 2, $\mathrm{E}^x (1 + d(X_{r_1})^2)^2 \le c_p^{(2)} (1 + d(x)^4)$. Using once again the estimates of Proposition 1 and Remark 1, we get

$$\mathrm{E}^x (F_r - F_0)^4 \le c_p^{(3)} \big(m(x)^2 r + r^2\big)\big(1 + d(x)^4\big).$$

Using this bound in (4.4), we obtain

$$\psi_2(x, r_1, r_2) \le c_p^{(4)} r_2 \big(m(x)^2 r_1 + r_1^2 (1 + d(x)^4)\big)^{1/2}.$$

Hence, from (4.3),

$$\mathrm{E}(F_b - F_a)^2 (F_b - F_c)^2 \le (c - b) \big[\big(F_a^2 (b - a) + (b - a)^2\big)\big(1 + d(X_a)^4\big)\big]^{1/2}.$$

Setting $a = \lfloor t_1 n \rfloor$, $b = \lfloor tn \rfloor$ and $c = \lfloor t_2 n \rfloor$, we get

$$\begin{aligned}
&\mathrm{E}\big(Z_n(t) - Z_n(t_1)\big)^2 \big(Z_n(t) - Z_n(t_2)\big)^2 \\
&\quad \le c_p^{(4)} n^{-2} (\lfloor t_2 n \rfloor - \lfloor tn \rfloor) \\
&\qquad \times \big(\mathrm{E}\big[m\big(X_{\lfloor t_1 n \rfloor}\big)^2 (\lfloor tn \rfloor - \lfloor t_1 n \rfloor) + (\lfloor tn \rfloor - \lfloor t_1 n \rfloor)^2\big]\big(1 + d\big(X_{\lfloor t_1 n \rfloor}\big)^4\big)\big)^{1/2}.
\end{aligned}$$

Using Corollary 2 and Proposition 1, one can see that

$$\mathrm{E} n^{-1} \big(m\big(X_{\lfloor t_1 n \rfloor}\big)^2 \big(1 + d\big(X_{\lfloor t_1 n \rfloor}\big)^4\big)\big)$$

is bounded from above by a constant since $t_1 \in [0, 1]$. We deduce that

$$\mathrm{E}\big(Z_n(t) - Z_n(t_1)\big)^2 \big(Z_n(t) - Z_n(t_2)\big)^2 \le c_p^{(4)} (t - t_1) \sqrt{t_2 - t}. \qquad \square$$

PROOF OF THEOREM 1. The convergence in law of $(Z_n)_{n \ge 0}$ to R is a consequence of the well-known Theorem 15.6 of [3], which states that convergence in law of the finite-dimensional distributions of a process, together with a tightness criterion, entails the convergence in law of the process on the Skorohod space. Propositions 2 and 3 provide the convergence of the finite-dimensional distributions of $Z_n(t)$ and the tightness criterion, respectively.

Furthermore, according to Lemma 2, the p-dimensional process

$$n^{-1/2}\big(X^{(i)}_{\lfloor n\cdot\rfloor} - X^{(1)}_{\lfloor n\cdot\rfloor}\big)_{1\le i\le p}$$

goes to 0 in probability as n goes to ∞ for the uniform distance on [0, 1] and hence for the Skorohod topology as well. $\square$

5. Large-deviations principle. We study the Laplace transform

$$\mathrm{E}\exp\big(tX_n^{(i)}\big).$$

Here again, the idea of the proof is to write recursion relations in which we approximate the transition probabilities of the selection steps by uniform ones. With no loss of generality, we consider the case of $X_n^{(1)}$. For the sake of brevity, we defer most of the proofs to Appendix B.

PROPOSITION 4. *For* $t \ge 0$,

$$\frac{1}{n}\log\mathrm{E}\exp\big(tX_n^{(1)}\big) \underset{n\to+\infty}{\longrightarrow} \log\cosh(t). \tag{5.1}$$

Let us sketch the idea of the proof. One has

$$\mathrm{E}\big(e^{t(X^{(1)}_{n+1})}\,|\,\mathcal{F}_n\big) = \mathbf{1}_{\{X_n\neq 0\}}\sum_{1\le u\le p}\frac{X_n^{(u)}}{S_n}e^{t(X_n^{(u)})}\cosh(t) + \mathbf{1}_{\{X_n=0\}}e^t\cosh(t).$$

Hence,

$$\mathrm{E}\big(e^{t(X^{(1)}_{n+1})}\big) = \cosh(t)\mathrm{E}\big(e^{t(X^{(1)}_{n})}\big)\big(1+m_n(t)\big), \tag{5.2}$$

with

$$\begin{aligned} m_n(t) = {}& \frac{1}{\mathrm{E}(e^{t(X_n^{(1)})})}\mathrm{E}\left(\mathbf{1}_{\{X_n\neq 0\}}\sum_{1\le u\le p}\frac{X_n^{(1)}-X_n^{(u)}}{S_n}e^{t(X_n^{(1)})}\right)\\ &+\frac{1}{\mathrm{E}(e^{t(X_n^{(1)})})}\mathrm{P}[X_n=0](e^t-1).\end{aligned} \tag{5.3}$$

We want to prove that $m_n(t)\to 0$ as $n\to\infty$. We first note that, in the definition of $m_n(t)$, the last term

$$\frac{1}{\mathrm{E}(e^{t(X_n^{(1)})})}\mathrm{P}[X_n=0](e^t-1)$$

goes to 0 since $\mathrm{P}[X_n=0](e^t-1)$ is bounded (t is fixed here!) and since the denominator $\mathrm{E}(e^{t(X_n^{(1)})})\to\infty$ as $n\to\infty$ (remember that $X_n^{(1)}\to\infty$ in probability). So we have to bound the first term. The inequality

$$\frac{|X_n^{(1)} - X_n^{(u)}|}{S_n} \le \frac{d(X_n)}{pF_n}$$

yields

$$\begin{aligned}\mathrm{E}\Biggl(\mathbf{1}_{\{X_n\neq 0\}} \sum_{1\le u\le p} \frac{X_n^{(1)} - X_n^{(u)}}{S_n} e^{t(X_n^{(1)})}\Biggr) \\ \le \mathrm{E}\biggl(\frac{d(X_n)}{pF_n}\mathbf{1}_{\{X_n\neq 0\}} \exp\Bigl(t \max_i X_n^{(i)}\Bigr)\biggr).\end{aligned}$$

Hence, Proposition 4 follows from the following result (stated as Lemma 8 in Appendix B): as $n \to \infty$,

$$\mathrm{E}\biggl(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}} \exp\Bigl(t \max_i X_n^{(i)}\Bigr)\biggr) = o\bigl(\mathrm{E}\bigl(\exp\bigl(tX_n^{(1)}\bigr)\bigr)\bigr).$$

The idea of the proof is the following: we condition the chain by its position at a time $n' \le n$, where n' is such that the probability that there is a crunch time between n' and n is high, and such that, on this event, the maximum possible diameter at time n is small, when compared to F_n. The key observation here, which we used already in the proof of Lemma 2, is that a nonnegligible piece of this event can be made independent of the mutation process.

6. Convergence of the cloud. For $p \in \mathbb{N}^*$, let

$$\mathbb{M}_p = \{(x_1, \dots, x_p) \in \mathbb{Z}^p;\ \forall i, j,\ x_i - x_j \in 2\mathbb{Z}\}$$

and let $\mathbb{H}_p$ be the quotient group $\mathbb{M}_p/\langle(1, \dots, 1)\rangle$: two elements of $\mathbb{M}_p$ are identified whenever they differ by an integer multiple of $(1, \dots, 1)$. Let $\Theta : \mathbb{M}_p \to \mathbb{H}_p$ denote the canonical projection.

Let $(\Xi_n)_{n\ge 0}$ be the selection–mutation chain on $\mathbb{Z}^p$ with uniform selection and mutation ± 1, starting from 0. Then $\xi = (\Theta(\Xi_n))_n$ is also a Markov chain on $\mathbb{H}_p$, irreducible, aperiodic and positively recurrent. Indeed,

$$\mathrm{P}(\xi_n = 0|\xi_{n-1}) \ge 2^{-p} p^{1-p} \qquad \text{a.s.}$$

Hence, ξ_n converges in law toward its invariant distribution π_p. We now prove Theorem 3, that is, that $\Theta(X_n)$ converges in law to π_p as well.

PROOF OF THEOREM 3. From Lévy's theorem (cf. [7]), it is enough to prove that

$$\begin{aligned}D_n(t) = \mathrm{E}\exp\bigl(i\bigl(t_1 X_n^{(1)} + \cdots + t_p X_n^{(p)}\bigr)\bigr) \\ - \mathrm{E}\exp\bigl(i\bigl(t_1 \Xi_n^{(1)} + \cdots + t_p \Xi_n^{(p)}\bigr)\bigr)\end{aligned}$$

goes to 0 for every p-tuple $t = (t_1, \dots, t_p)$ of real numbers summing to 0. We now consider the effects of mutation and selection over D_n. The mutation steps of X and Ξ multiply $D_n(t)$ by $\cos(t_1) \times \cdots \times \cos(t_p)$; hence, taking into account the selection steps as well, one gets

$$\begin{aligned}
D_{n+1}(t) = {}& \mathrm{E}\mathbf{1}_{\{X_n=0\}} e^{i(t_1+\cdots+t_p)} \cos(t_1) \times \cdots \times \cos(t_p) \\
& + \mathrm{E} \sum_{1 \le j_1, \dots, j_p \le p} \cos(t_1) \times \cdots \times \cos(t_p) \\
& \times \Biggl[\mathbf{1}_{\{X_n \neq 0\}} \Biggl(\prod_{k=1}^p \frac{X_n^{(j_k)}}{S_n} \Biggr) \exp\Biggl(i \sum_{k=1}^p t_k X_n^{(j_k)} \Biggr) \\
& \qquad - \frac{1}{p^p} \exp\Biggl(i \sum_{k=1}^p t_k \Xi_n^{(j_k)} \Biggr) \Biggr].
\end{aligned}$$

To use the approximation of the selection transition probabilities of X by those of a uniform selection, we add and subtract p^{-p} to $\prod_k X_n^{(j_k)}/S_n$:

$$\begin{aligned}
D_{n+1}(t) = {}& \mathrm{E}\mathbf{1}_{\{X_n=0\}} e^{i(t_1+\cdots+t_p)} \cos(t_1) \times \cdots \times \cos(t_p) \\
& + \mathrm{E} \sum_{1 \le j_1, \dots, j_p \le p} \cos(t_1) \times \cdots \times \cos(t_p) \mathbf{1}_{\{X_n \neq 0\}} \\
& \times \Biggl(\Biggl(\prod_{k=1}^p \frac{X_n^{(j_k)}}{S_n} \Biggr) - \frac{1}{p^p} \Biggr) \exp\Biggl(i \sum_{k=1}^p t_k X_n^{(j_k)} \Biggr) \\
& + \mathrm{E} \sum_{1 \le j_1, \dots, j_p \le p} \cos(t_1) \times \cdots \times \cos(t_p) \\
& \times \frac{1}{p^p} \Biggl[\mathbf{1}_{\{X_n \neq 0\}} \exp\Biggl(i \sum_{k=1}^p t_k X_n^{(j_k)} \Biggr) - \exp\Biggl(i \sum_{k=1}^p t_k \Xi_n^{(j_k)} \Biggr) \Biggr].
\end{aligned}$$

Replacing $\mathbf{1}_{\{X_n \neq 0\}}$ by $1 - \mathbf{1}_{\{X_n=0\}}$, we finally obtain

$$\begin{aligned}
D_{n+1}(t) = {}& \mathrm{E}\mathbf{1}_{\{X_n=0\}} e^{i(t_1+\cdots+t_p)} \cos(t_1) \times \cdots \times \cos(t_p) \\
& + \mathrm{E} \sum_{1 \le j_1, \dots, j_p \le p} \cos(t_1) \times \cdots \times \cos(t_p) \mathbf{1}_{\{X_n \neq 0\}} \\
& \times \Biggl(\Biggl(\prod_{k=1}^p \frac{X_n^{(j_k)}}{S_n} \Biggr) - \frac{1}{p^p} \Biggr) \exp\Biggl(i \sum_{k=1}^p t_k X_n^{(j_k)} \Biggr)
\end{aligned}$$

$$
\begin{aligned}
&- \mathrm{E} \sum_{1 \le j_1, \dots, j_p \le p} \cos(t_1) \times \cdots \times \cos(t_p) \\
&\quad \times \frac{1}{p^p} \mathbf{1}_{\{X_n = 0\}} \exp\left(i \sum_{k=1}^{p} t_k X_n^{(j_k)}\right) \\
&+ \mathrm{E} \sum_{1 \le j_1, \dots, j_p \le p} \cos(t_1) \times \cdots \times \cos(t_p) \\
&\quad \times \frac{1}{p^p} \left[\exp\left(i \sum_{k=1}^{p} t_k X_n^{(j_k)}\right) - \exp\left(i \sum_{k=1}^{p} t_k \Xi_n^{(j_k)}\right)\right].
\end{aligned}
$$

DEFINITION 3. J_p is the set of maps $\{1, \dots, p\} \to \{1, \dots, p\}$. For $f \in J_p$, we note $\# f = \#\{f(i), i \in \{1, \dots, p\}\}$. Let $J_p^0 = \{f \in J_p, \# f = 1\}$. For $f \in J_p$ and $t \in \mathbb{R}^p$, we define $f(t) \in \mathbb{R}^p$ by

$$
(f(t))_i = \sum_{f(j)=i} t_j \qquad \forall\, i \in \{1, \dots, p\}.
$$

We note that $(f \circ g)(t) = f(g(t))$. With this notation, the last expression of $D_{n+1}(t)$ leads to

$$
D_{n+1}(t) = \varepsilon_n(t) + \frac{1}{p^p} \sum_{f \in J_p} \cos(t_1) \times \cdots \times \cos(t_p)\, D_n(f(t)),
$$

where, owing to Lemma 4,

$$
\begin{aligned}
|\varepsilon_n(t)| \le \varepsilon_n &= 2\mathrm{P}[X_n = 0] + p^p \mathrm{E} \mathbf{1}_{\{X_n \neq 0\}} d(X_n)/F_n \\
&\le 2\mathrm{P}[X_n = 0] + p^p \bigl(\mathrm{E} d(X_n)^2\bigr)^{1/2} \bigl(\mathrm{E} \mathbf{1}_{\{X_n \neq 0\}} F_n^{-2}\bigr)^{1/2}.
\end{aligned}
$$

Hence, owing to Corollary 1 and Lemma 2, $\varepsilon_n \to 0$ as $n \to \infty$. Since the numbers t_i add to 0, $D_n(f(t)) = 0$ if $f \in J_p^0$. Hence,

$$
D_{n+1}(t) = \varepsilon_n(t) + \frac{1}{p^p} \sum_{f \in J_p \setminus J_p^0} \cos(t_1) \times \cdots \times \cos(t_p) D_n(f(t)). \tag{6.1}
$$

The set

$$
\mathcal{J}_p(t) = \{ f_k \circ \cdots \circ f_1(t),\ k \in \mathbb{N},\ f_1, \dots, f_k \in J_p \}
$$

is finite, and, for all $u \in \mathcal{J}_p(t)$, $\sum_i u_i = 0$. Introducing

$$
\bar{D}_n(t) = \sup_{u \in \mathcal{J}_p(t)} |D_n(u)|,
$$

(6.1) leads to

$$\bar{D}_{n+1}(t) \le \varepsilon_n + \frac{|J_p \setminus J_p^0|}{p^p} \bar{D}_n(t).$$

Since $|J_p \setminus J_p^0|/p^p < 1$, this proves that $D_n(t)$ goes to 0. $\square$

Although we do not have a close formula for π_p, its Fourier transform can be algorithmically computed as follows. Let

$$\phi(t) = \mathrm{E}\exp\left(i \sum_{j=1}^{p} t_j x(j)\right),$$

where $t = (t_1, \dots, t_p)$ is a p-tuple of real numbers summing to 0, and let $x = (x(1), \dots, x(p)) \in \mathbb{M}_p$ be a random vector, such that $\Theta(x)$ has the law π_p [one can check that $\exp(i \sum_{j=1}^{p} t_j x(j))$ is really a function of $\Theta(x)$]. The invariance of π_p with respect to Ξ yields

$$\begin{aligned}\phi(t) &= \cos(t_1) \times \cdots \times \cos(t_p) p^{-p} \sum_{1 \le j_1, \dots, j_p \le p} \mathrm{E}\exp\left(i \sum_{l=1}^{p} t_l x(j_l)\right) \\ &= \cos(t_1) \times \cdots \times \cos(t_p) p^{-p} \sum_{f \in J_p} \phi(f(t)).\end{aligned} \tag{6.2}$$

Note $\#t = \#\{i,\ t_i \neq 0\}$. The set $\{f(t),\ f \in J_p\}$ contains p-tuples u such that either $\phi(u) = \phi(t)$ (when $\#u = \#t$, i.e., when u is deduced from t by a permutation) or $\#u < \#t$. The relation (6.2) gives a method for computing $\phi(t)$ by induction on $\#t$, since $\phi(t) = 1$ when $t = 0$. As an example, we compute $\phi(t)$ when $\#t = 2$.

PROPOSITION 5. *For all $1 \le i \neq j \le p$, the sequence $X_n^{(i)} - X_n^{(j)}$ converges in law to a random variable Δ of law*

$$(1-\alpha)\delta_0 + \sum_{n \ge 1} (1-a)^{n-1} a (\delta_{2n} + \delta_{-2n}) \alpha/2,$$

where $\alpha = \sqrt{p}/(\sqrt{p}+1)$ and $a = 2/(\sqrt{p}+1)$. The variable Δ may be written as $\Delta = 2ZG$, where the law of the random variable Z is

$$(\delta_1 + \delta_{-1})\alpha/2 + (1-\alpha)\delta_0$$

and G is a geometric random variable independent of Z: for $n \ge 1$,

$$\mathrm{P}[G = n] = (1-a)^{n-1} a.$$

PROOF. For $t \in \mathbb{R}$, (6.2) gives

$$\begin{aligned}
&\phi(t, -t, 0, \dots, 0) \\
&\quad = \cos^2 t \times p^{-p}\left(\sum_{f \in J_p, f(1) \neq f(2)} \phi(t, -t, 0, \dots, 0) + \sum_{f \in J_p, f(1) = f(2)} 1\right) \\
&\quad = \cos^2 t\left(\frac{p-1}{p}\phi(t, -t, 0, \dots, 0) + \frac{1}{p}\right),
\end{aligned}$$

leading to

$$\phi(t, -t, 0, \dots, 0) = \frac{\cos^2 t}{p - (p-1)\cos^2 t}.$$

The result follows by a simple computation. □

7. Simulations. We have performed numerical simulations of the algorithms studied in this paper. These simulations first yielded empirical insights into the behavior of the chain. Afterwards, they showed that the asymptotic results of the previous sections were indeed observed on real finite-size systems.

As regards the technical side of the simulations, we designed a C computer program using a Mersenne–Twister random generator; see [11]. We used an egcs C compiler on a generic Pentium II® computer.

Figure 1 shows the time evolution of the average fitness F_n for a single realization of the chain. Figure 2 shows the average time evolution of F_n performed over 10^4 independent realizations for several values of the population size p.

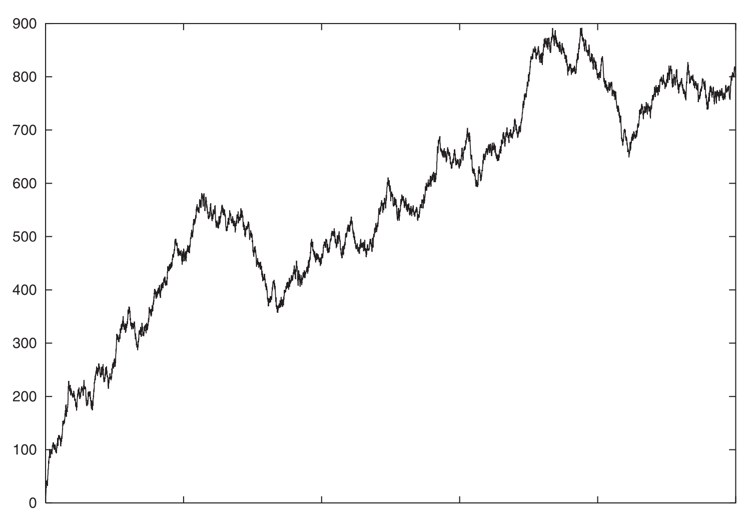

FIG. 1. *Single realization of F_n up to time 10^5, 4 particles.*

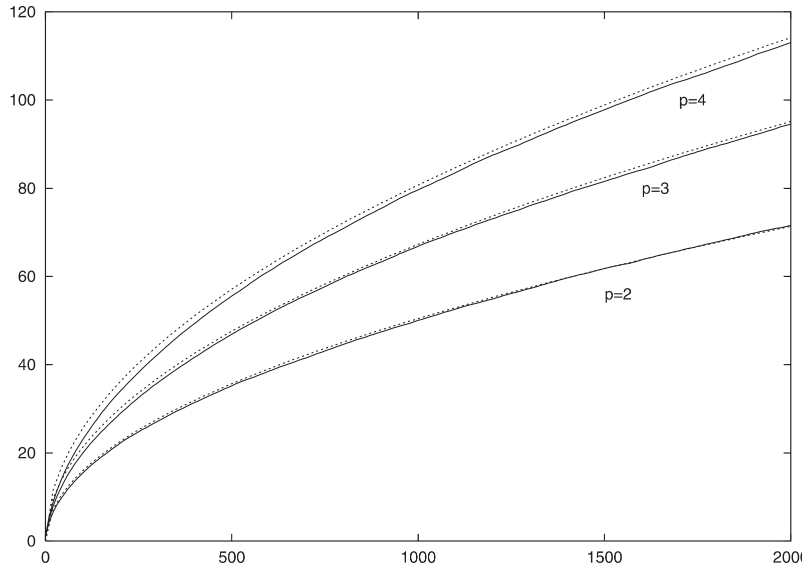

FIG. 2. *Average of F_n over 10^4 realizations up to time 10^5, $p = 2, 3, 4$ particles.*

Figure 3 displays the empirical distribution of $X_n^{(1)}/\sqrt{n}$ for $n = 2 \cdot 10^4$ and $p = 2$. Figure 4 displays the empirical joint distribution of $(X_n^{(2)} - X_n^{(1)}, X_n^{(3)} - X_n^{(1)})$ for $n = 5 \cdot 10^4$ and $p = 3$. Dashed lines display the theoretical asymptotical values.

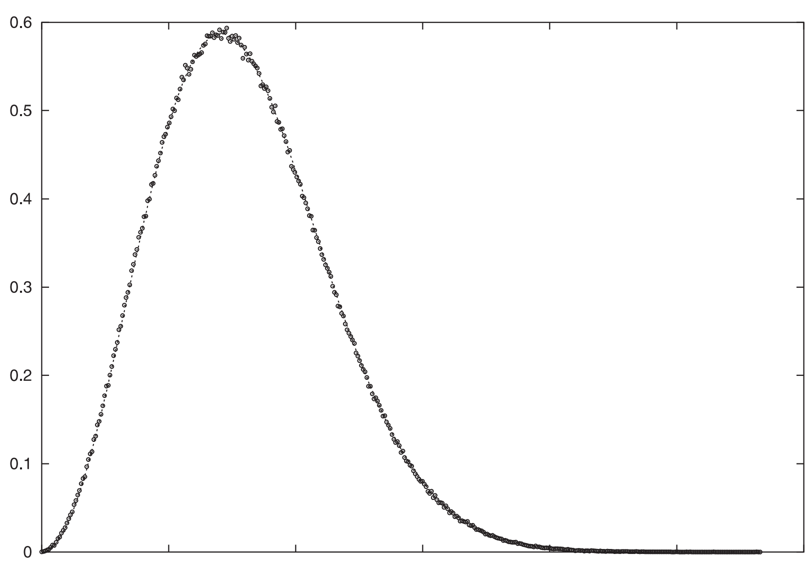

FIG. 3. *Empirical distribution of $X_n^{(1)}/\sqrt{n}$, 2 particles, time 2×10^4, 2×10^6 realizations.*

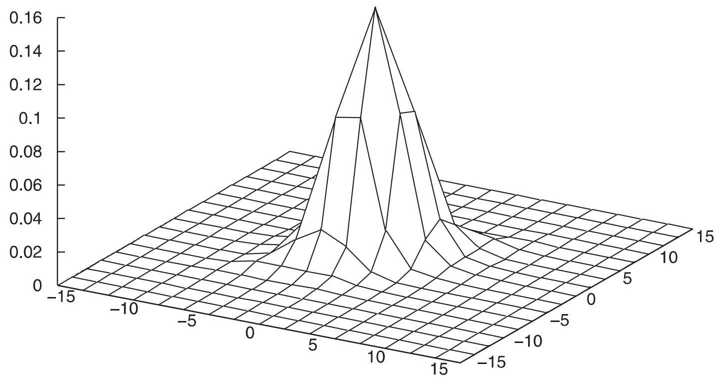

FIG. 4. *Frequency of the relative positions of the three particles*, *time* 5×10^4, 10^6 *realizations*.

APPENDIX A

PROOF OF PROPOSITION 1. By induction on h, (P_0) holds trivially with $f(x,n,0)=1$. Suppose that (P_i) holds for all $i \le h-1$, with $h \ge 1$. Then

$$\mathrm{E}^x F_n^{2h} = m(x)^{2h} + \mathrm{E}^x \sum_{r=0}^{n-1} (F_{r+1}^{2h} - F_r^{2h}). \tag{A.1}$$

Expanding, we get

$$\begin{aligned}\mathrm{E}^x(F_{r+1}^{2h} - F_r^{2h}) &= \mathrm{E}^x\big((F_r + \Delta F_r)^{2h} - F_r^{2h}\big)\\ &= \binom{2h}{1}\mathrm{E}^x\big(F_r^{2h-1}\mathrm{E}^x(\Delta F_r|\mathcal{F}_r)\big)\\ &\quad + \binom{2h}{2}\mathrm{E}^x\big(F_r^{2h-2}\mathrm{E}^x\big((\Delta F_r)^2\big|\mathcal{F}_r\big)\big)\\ &\quad + \sum_{l=3}^{2h}\binom{2h}{l}\mathrm{E}^x\big(F_r^{2h-l}(\Delta F_r)^l\big).\end{aligned} \tag{A.2}$$

We start with the contribution of the rightmost sum in (A.2): for $3 \le l \le 2h$,

$$\big|\mathrm{E}^x F_r^{2h-l}(\Delta F_r)^l\big| \le \sum_{r=0}^{n-1} \mathrm{E}^x (1+F_r)^{2h-l}|\Delta F_r|^l. \tag{A.3}$$

Since $(1+F_r)^{2h-l} = (1+F_r)^{2h-l+1/2} \times (1+F_r)^{-1/2}$, Hölder's inequality yields

$$\begin{aligned}\big|\mathrm{E}^x F_r^{2h-l}(\Delta F_r)^l\big| &\le \Big(\mathrm{E}^x\big((1+F_r)^{2h-l+1/2}\big)^{\theta}\Big)^{1/\theta}\\ &\quad \times \Big(\mathrm{E}^x\big((1+F_r)^{-1/2}|\Delta F_r|\big)^{l\theta'}\Big)^{1/\theta'},\end{aligned}$$

where $\theta = (2h-2)/(2h-l+1/2) > 1$ and $\theta^{-1} + \theta'^{-1} = 1$. Since $l \geq 3$ and $|\Delta F_r| \leq d(F_r) + 2$,

$$\begin{aligned} &\bigl|\mathrm{E}^x F_r^{2h-l} (\Delta F_r)^l\bigr| \\ &\qquad \leq \bigl(\mathrm{E}^x (1+F_r)^{2h-2}\bigr)^{1/\theta} \Bigl(\mathrm{E}^x \bigl((1+F_r)^{-1/2} \bigl(d(X_r)+2\bigr)\bigr)^{l\theta'}\Bigr)^{1/\theta'}. \end{aligned} \tag{A.4}$$

We first bound the rightmost factor in (A.4): from Hölder's inequality,

$$\begin{aligned} &\mathrm{E}^x \bigl((1+F_r)^{-1/2} \bigl(d(X_r)+2\bigr)\bigr)^{l\theta'} \\ &\qquad \leq \Bigl[\mathrm{E}^x \bigl((1+F_r)^{-1/2}\bigr)^{2l\theta'} \mathrm{E}^x \bigl(d(X_r)+2\bigr)^{2l\theta'}\Bigr]^{1/2} \\ &\qquad \leq \bigl[\mathrm{E}^x (1+F_r)^{-l\theta'}\bigr]^{1/2} c_{p,h}^{(5)} \bigl(1+d(x)^{8lh}\bigr)^{1/2}, \end{aligned}$$

according to Lemma 2 and since $2l\theta' \leq 8lh$. Furthermore,

$$\mathrm{E}^x (1+F_r)^{-l\theta'} \leq \mathrm{E}\bigl(1+m(Y_r)\bigr)^{-l\theta'},$$

where (Y_n) is the process defined in Lemma 1. This last term goes to 0 as n goes to ∞ (and does not depend on x).

We now bound the leftmost factor in (A.4):

$$\bigl(\mathrm{E}^x (1+F_r)^{2h-2}\bigr)^{1/\theta} \leq \mathrm{E}^x (1+F_r)^{2h-2} = \sum_{i=0}^{2h-2} \binom{2h-2}{i} \mathrm{E}^x F_r^i.$$

Using the induction hypothesis $(P_0, \dots, P_{h-1})$, we see that the last term of (A.2) satisfies (Q_{h-1}). Hence, its contribution in (A.1)—summing for $r = 0, \dots, n-1$—satisfies (Q_h).

Now, we study the first term in (A.2). From Lemma 3,

$$\begin{aligned} &\binom{2h}{1} \mathrm{E}^x \bigl(F_r^{2h-1} \mathrm{E}^x (\Delta F_r | \mathcal{F}_r)\bigr) \\ &\qquad = 2hp^{-2} \mathrm{E}^x \sum_{1 \leq i < j \leq p} F_r^{2h-2} \bigl(X_r^{(i)} - X_r^{(j)}\bigr)^2. \end{aligned} \tag{A.5}$$

For $0 \leq s \leq r-1$ and $i \neq j$,

$$\mathrm{E}^x F_{s+1}^{2h-2} \bigl(X_{s+1}^{(i)} - X_{s+1}^{(j)}\bigr)^2 = \mathrm{E}^x (F_s + \Delta F_s)^{2h-2} \bigl(X_{s+1}^{(i)} - X_{s+1}^{(j)}\bigr)^2,$$

and, expanding $(F_s + \Delta F_s)^{2h-2}$, we get

$$
\begin{aligned}
&\mathrm{E}^x F_{s+1}^{2h-2}(X_{s+1}^{(i)} - X_{s+1}^{(j)})^2 \\
&\quad = \sum_{l=0}^{2h-2} \binom{2h-2}{l} \mathrm{E}^x F_s^{2h-2-l}(\Delta F_s)^l (X_{s+1}^{(i)} - X_{s+1}^{(j)})^2 \\
&\quad = \mathrm{E}^x F_s^{2h-2} \mathrm{E}^x \big((X_{s+1}^{(i)} - X_{s+1}^{(j)})^2 | \mathcal{F}_s\big) \\
&\qquad + \underbrace{\sum_{l=1}^{2h-2} \binom{2h-2}{l} \mathrm{E}^x F_s^{2h-2-l}(\Delta F_s)^l (X_{s+1}^{(i)} - X_{s+1}^{(j)})^2}_{(\star)}.
\end{aligned}
\tag{A.6}
$$

Using the same method as for the last term in (A.2), we show that the sum $(\star)$ satisfies (Q_{h-1}).

Moreover, conditioning by $\mathcal{F}_s$ yields

$$
\begin{aligned}
&\mathrm{E}^x \big((X_{s+1}^{(i)} - X_{s+1}^{(j)})^2 | \mathcal{F}_s\big) \\
&\qquad = 2 + \mathrm{E}^x \big((Y_s^{(i)} - Y_s^{(j)})^2 | \mathcal{F}_s\big) \\
&\qquad = 2 + \mathbf{1}_{\{X_s \neq 0\}} \sum_{1 \le a, b \le p} \frac{X_s^{(a)} X_s^{(b)}}{S_s^2} (X_s^{(a)} - X_s^{(b)})^2.
\end{aligned}
\tag{A.7}
$$

If we replace in (A.7) the selection transition probabilities $X_s^{(a)} X_s^{(b)} / S_s^2$ by uniform selection transition probabilities p^{-2}, we get in (A.6) an error term that we bound using Lemma 4: for all $1 \le a, b \le p$,

$$
\begin{aligned}
&\mathbf{1}_{\{X_s \neq 0\}} \left| F_s^{2h-2} \left(\frac{X_s^{(a)} X_s^{(b)}}{S_s^2} - \frac{1}{p^2} \right) (X_s^{(a)} - X_s^{(b)})^2 \right| \\
&\qquad \le 2 p^{-1} d(X_s) F_s^{2h-3}.
\end{aligned}
$$

Hence, for $0 \le s \le r - 1$, (A.6) can be written as

$$
\begin{aligned}
\mathrm{E}^x F_{s+1}^{2h-2}(X_{s+1}^{(i)} - X_{s+1}^{(j)})^2 = 2\mathrm{E}^x F_s^{2h-2} + \frac{p-1}{p} \mathrm{E}^x F_s^{2h-2} (X_s^{(i)} - X_s^{(j)})^2 \\
+ \beta_1(x, s, h-1),
\end{aligned}
$$

where $\beta_1(x, s, h-1)$ satisfies (Q_{h-1}). Summing these estimates for $s = 0, \ldots,$

$r-1$, we get, for all $0 \le r \le n-1$,

$$\begin{aligned}
\mathrm{E}^x F_r^{2h-2}\big(X_r^{(i)} - X_r^{(j)}\big)^2 &= \left(\frac{p-1}{p}\right)^r (x_i - x_j)^2 m(x)^{2h-2} \\
&\quad + \sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} \big(2\mathrm{E}^x F_k^{2h-2} + \beta_1(x,k,h-1)\big) \\
&= \left(\frac{p-1}{p}\right)^r (x_i - x_j)^2 m(x)^{2h-2} \\
&\quad + \beta_2(x,r,h-1) + \sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} 2\mathrm{E}^x F_k^{2h-2},
\end{aligned}$$

where $\beta_2(x,r,h-1)$ satisfies (Q_{h-1}). Hence, according to (P_{h-1}),

$$\begin{aligned}
&\sum_{r=0}^{n-1} \mathrm{E}^x F_r^{2h-2}\big(X_r^{(i)} - X_r^{(j)}\big)^2 \\
&\qquad = p\left(1 - \left(\frac{p-1}{p}\right)^n\right)(x_i - x_j)^2 m(x)^{2h-2} \\
&\qquad\quad + \sum_{r=0}^{n} \beta_2(x,r,h-1) \\
&\qquad\quad + \sum_{r=0}^{n-1}\sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} 2\big(f(x,k,h-1) + \alpha(x,k,h-1)\big).
\end{aligned}$$

Here, the two first lines and the contribution of $\alpha(x,k,h-1)$ satisfy (Q_h). It remains to show that

$$\begin{aligned}
&\sum_{r=0}^{n-1} \mathrm{E}^x F_r^{2h-2}\big(X_r^{(i)} - X_r^{(j)}\big)^2 \\
&\qquad = 2\sum_{r=0}^{n-1}\sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1) + \beta_3(x,n,h),
\end{aligned} \tag{A.8}$$

where $\beta_3(x,r,h)$ satisfies (Q_h). We now consider the sum over k in (A.8). Using (P_{h-1}),

$$\begin{aligned}
&\sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1) \\
&\qquad = \sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} \sum_{i=0}^{h-1} f_i(h-1) m(x)^{2i} k^{h-i} \\
&\qquad = \sum_{i=0}^{h-1} f_i(h-1) m(x)^{2i} \sum_{k=0}^{r-1} \left(\frac{p-1}{p}\right)^{r-1-k} k^{h-i}.
\end{aligned}$$

From Lemma 5, one has

$$\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1) = p\sum_{i=0}^{h-1} f_i(h-1)m(x)^{2i} r^{h-i}[1+\epsilon_i(r)],$$

where the ϵ_i's are sequences such that $\epsilon_i(n)$ goes to 0 as n goes to ∞. Using again (P_{h-1}), we get

$$\sum_{k=0}^{r-1}\left(\frac{p-1}{p}\right)^{r-1-k} f(x,k,h-1) = pf(x,r,h-1) + \beta_4(x,r,h-1),$$

where $\beta_4(x,r,h-1)$ satisfies (Q_{h-1}). Plugging this estimate into (A.8) gives

$$\sum_{r=0}^{n-1} \mathrm{E}^x F_r^{2h-2}\big(X_r^{(i)} - X_r^{(j)}\big)^2 = \beta_5(x,n,h) + 2p\sum_{r=0}^{n-1} f(x,r,h-1), \tag{A.9}$$

where $\beta_5(x,r,h)$ satisfies (Q_h).

Plugging (A.9) into (A.5) gives

$$\begin{aligned}\sum_{r=0}^{n}\binom{2h}{1}\mathrm{E}^x F_r^{2h-1}\mathrm{E}^x(\Delta F_r|\mathcal{F}_r) &= \beta_6(x,n,h) + 2h/p^2 \sum_{1\le i<j\le p} 2p\sum_{r=0}^{n-1} f(x,r,h-1)\\ &= \beta_7(x,n,h) + 2h(p-1)\sum_{r=0}^{n-1} f(x,r,h-1),\end{aligned} \tag{A.10}$$

where $\beta_7(x,n,h)$ satisfies (Q_h).

We now study the second term in (A.2). From Lemma 3, using Lemma 4 to approximate the selection transition probabilities by uniform ones, as we did for the first term of (A.2), we get

$$\begin{aligned}\mathrm{E}^x F_r^{2h-2}\mathrm{E}^x\big((\Delta F_r)^2|\mathcal{F}_r\big) &= p^{-1}\mathrm{E}^x F_r^{2h-2}\\ &\quad + p^{-4}\sum_{1\le i\ne j\le p}\sum_{1\le a,b\le p} \mathrm{E}^x F_r^{2h-2}\big(X_r^{(a)} - X_r^{(i)}\big)\big(X_r^{(b)} - X_r^{(j)}\big)\\ &\quad + p^{-3}\sum_{1\le i\le p}\sum_{1\le a\le p} \mathrm{E}^x F_r^{2h-2}\big(X_r^{(a)} - X_r^{(i)}\big)^2\\ &\quad + \beta_8(x,r,h-1),\end{aligned} \tag{A.11}$$

where $\beta_8(x,r,h-1)$ satisfies (Q_{h-1}).

For an arbitrary sequence $(x_i)_{1\le i\le p}$, one has

$$\sum_{1\le i\neq j\le p}\sum_{1\le a,b\le p}(x_a-x_i)(x_b-x_j)$$
$$=\sum_{1\le a,b,i,j\le p}(x_a-x_i)(x_b-x_j)-\sum_{i=j,a,b}(x_a-x_i)(x_b-x_j).$$

By symmetry, the first sum on the right-hand side is 0. As for the last sum, we separate cases $a=b$ and $a\neq b$:

$$\sum_{1\le i\neq j\le p}\sum_{1\le a,b\le p}(x_a-x_i)(x_b-x_j)$$
$$=-\sum_{i,a}(x_a-x_i)^2-\sum_{i,a\neq b}(x_a-x_i)(x_b-x_i).$$

In the last sum, let (u,v,w) be the ordered triple with elements i,a,b. With this notation, the last equation reads

$$\sum_{1\le i\neq j\le p}\sum_{1\le a,b\le p}(x_a-x_i)(x_b-x_j)$$
$$=-\sum_{i,a}(x_a-x_i)^2$$
$$-2\sum_{1\le u<v<w\le p}(x_v-x_u)(x_w-x_u)+(x_u-x_v)(x_w-x_v)$$
$$+(x_u-x_w)(x_v-x_w).$$

Note that, for all x,y,z,

$$2\big[(y-x)(z-x)+(x-y)(z-y)+(x-z)(y-z)\big]$$
$$=(x-y)^2+(y-z)^2+(x-z)^2,$$

so that

$$\sum_{1\le i\neq j\le p}\sum_{1\le a,b\le p}(x_a-x_i)(x_b-x_j)$$
$$=-\sum_{i,a}(x_a-x_i)^2-\sum_{1\le u<v<w\le p}(x_u-x_v)^2+(x_v-x_w)^2+(x_u-x_w)^2.$$

Applying this to the sequence $(X_r^{(k)})_k$ and plugging it and (A.9) into (A.11), we get

$$\sum_{r=0}^{n-1}\mathrm{E}^x F_r^{2h-2}\mathrm{E}^x\big((\Delta F_r)^2\big|\mathcal{F}_r\big)$$
$$=\left[p^{-1}-p^{-4}\left(p(p-1)+3\binom{p}{3}\right)2p+p^{-3}p(p-1)2p\right] \tag{A.12}$$

$$\times \sum_{r=0}^{n-1} f(x,r,h-1) + \beta_9(x,n,h)$$

$$= \beta_9(x,n,h) + \sum_{r=0}^{n-1} f(x,r,h-1),$$

where $\beta_9(x,n,h)$ satisfies (Q_h).

Finally, owing to (A.2), (A.10) and (A.12),

$$\mathrm{E}^x F_n^{2h} = m(x)^{2h} + [2h(p-1) + h(2h-1)] \sum_{r=0}^{n-1} f(x,r,h-1) + \beta_{10}(x,n,h),$$

where $\beta_{10}(x,n,h)$ satisfies (Q_h). Hence, (P_h) holds, with

$$\begin{aligned} & f_h(h) = 1, \\ \text{(A.13)} \qquad & f_i(h) = \frac{2h(p-1) + h(2h-1)}{h-i} f_i(h-1), \qquad 0 \le i \le h-1. \end{aligned}$$ □

APPENDIX B

Lemma 7 means that $\mathrm{E}(e^{t(X_n^{(1)})})$ grows from step to step at least as fast as it should, according to Proposition 4, which we want to prove. Corollary 3 states a similar result for $\mathrm{E}(e^{t \max_i X_n^{(i)}})$.

LEMMA 7. *For all $n \ge 0$ and $t \ge 0$, $m_n(t) \ge 0$.*

PROOF. Obviously, the last term in (5.3) is nonnegative since $t \ge 0$. As regards the first term, by symmetry,

$$\begin{aligned} & 2\mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} (X_n^{(1)} - X_n^{(u)}) e^{t(X_n^{(1)})} \\ & \quad = \mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} (X_n^{(1)} - X_n^{(u)}) e^{t(X_n^{(1)})} \\ & \qquad + \mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} (X_n^{(u)} - X_n^{(1)}) e^{t(X_n^{(u)})} \\ & \quad = \mathrm{E}\mathbf{1}_{\{X_n \neq 0\}} S_n^{-1} (X_n^{(1)} - X_n^{(u)}) \big(e^{t(X_n^{(1)})} - e^{t(X_n^{(u)})}\big), \end{aligned}$$

which is nonnegative since $y \mapsto e^y$ is an increasing function. □

COROLLARY 3. *For all $t \ge 0$ and $n \ge q$,*

$$\mathrm{E}\Big(\exp\Big(t \max_i X_{n-q}^{(i)}\Big)\Big) \le p(\cosh(t))^{-q} \mathrm{E}\Big(\exp\Big(t \max_i X_n^{(i)}\Big)\Big).$$

PROOF. By Lemma 7,

$$\mathrm{E}\exp\left(tX_n^{(1)}\right) \geq \cosh^q(t)\mathrm{E}\exp\left(tX_{n-q}^{(1)}\right).$$

Now,

$$\begin{aligned}
\mathrm{E}\Big(\exp\Big(t\max_i X_{n-q}^{(i)}\Big)\Big) &\leq \sum_{1\leq i\leq p} \mathrm{E}\exp\left(tX_{n-q}^{(i)}\right) \\
&\leq p\mathrm{E}\big(\exp\left(tX_{n-q}^{(1)}\right)\big) \\
&\leq p(\cosh(t))^{-q}\mathrm{E}\big(\exp\left(tX_n^{(1)}\right)\big) \\
&\leq p(\cosh(t))^{-q}\mathrm{E}\Big(\exp\Big(t\max_i X_n^{(i)}\Big)\Big). \qquad \square
\end{aligned}$$

We now prove the key result.

LEMMA 8. *As n goes to ∞,*

$$\mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp\Big(t\max_i X_n^{(i)}\Big)\right) = o\big(\mathrm{E}\big(\exp\left(tX_n^{(1)}\right)\big)\big). \tag{B.1}$$

PROOF. Let $(l_n)_n$ and $(u_n)_n$ be two increasing sequences of integers. We first show that only large values of F_n contribute to the left-hand side of (B.1).

Using the upper bound $d(X_n) \leq pF_n$,

$$\begin{aligned}
&\mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp\Big(t\max_i X_n^{(i)}\Big)\mathbf{1}_{\{\max_i X_n^{(i)}\leq u_n\}}\right) \\
&\quad\leq \mathrm{E}\Big(p\exp(tu_n)\mathbf{1}_{\{\max_i X_n^{(i)}\leq u_n\}}\Big) \\
&\quad\leq p\exp(tu_n).
\end{aligned} \tag{B.2}$$

Assume that the selection steps are built through the sequence of random parameters U_m as defined in the proof of Lemma 2. Let

$$Q_n = \bigcup_{m=n-l_n}^{n-1}\{U_m \leq p^{1-p}\},$$

so that, on Q_n, $d(X_n) \leq 2l_n$ since there is a crunch time between $n-l_n$ and n. Hence,

$$\begin{aligned}
&\mathrm{E}\left(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp\Big(t\max_i X_n^{(i)}\Big)\mathbf{1}_{\{\max_i X_n^{(i)}\geq u_n\}}\mathbf{1}_{Q_n}\right) \\
&\quad\leq \frac{2l_n}{u_n/p}\mathrm{E}\Big(\exp\Big(t\max_i X_n^{(i)}\Big)\mathbf{1}_{Q_n}\Big) \\
&\quad\leq \frac{2pl_n}{u_n}\sum_{i=1}^{p}\mathrm{E}\big(\exp\left(tX_n^{(i)}\right)\mathbf{1}_{Q_n}\big).
\end{aligned} \tag{B.3}$$

We introduce additional notation concerning the one-step transitions of (X_n).

- *Selection step*: $X_m \to Y_m$. $Y_m^{(i)} = X_m^{(f_m(i))}$, where the f_m's are random functions on $\{1, \dots, p\}$ such that, knowing $\mathcal{F}_m$, the $(f_m(i))_{1 \le i \le p}$ are i.i.d. with law
$$\mathbf{1}_{\{X_n \neq 0\}} S_m^{-1} \sum_{j=1}^{p} X_m^{(j)} \delta_j + \mathbf{1}_{\{X_n = 0\}} \delta_{(1,\dots,1)}.$$
Obviously, f_m and the random parameter U_m introduced above are not independent from each other.
- *Mutation step*: $Y_m \to X_{m+1}$. $X_{m+1} = Y_m + \eta_m$, where $\eta_m = (\eta_m^{(1)}, \dots, \eta_m^{(p)})$ and the random variables $\eta_m^{(i)} \in \{-1, +1\}$, $1 \le i \le p$, are independent from each other and independent from $\mathcal{G}_m$, with symmetric Bernoulli law.

With this notation, let, for $n - l_n \le m \le n$,
$$g_m = f_n \circ \cdots \circ f_m.$$
Thus, for all $j \in \{1, \dots, p\}$,
$$\begin{aligned}
&\mathrm{E}\big(\exp\big(tX_n^{(j)}\big)\mathbf{1}_{Q_n} \big| \mathcal{F}_{n-l_n}\big) \\
&\quad = \mathrm{E}\left(\exp\left(t\left[X_{n-l_n}^{(g_{n-l_n}(j))} + \sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}\right]\right)\mathbf{1}_{Q_n} \middle| \mathcal{F}_{n-l_n}\right) \\
&\quad \le \exp\left(t \max_i X_{n-l_n}^{(i)}\right) \\
&\qquad \times \mathrm{E}\left(\exp\left(t \sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}\right)\mathbf{1}_{Q_n} \middle| \mathcal{F}_{n-l_n}\right).
\end{aligned} \tag{B.4}$$
Knowing $(U_{n-l_n}, \dots, U_{n-1})$, g_m depends only on η_r for $r < m$. Since the η_m's are i.i.d. Bernoulli random vectors independent of the U_m's, the conditional law, knowing $(U_{n-l_n}, \dots, U_{n-1})$, of
$$\sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}$$
is the law of a sum of l_n symmetric Bernoulli ± 1 independent random variables. Hence,
$$\mathrm{E}\left(\exp\left(t \sum_{m=n-l_n}^{n-1} \eta_m^{(g_m(j))}\right)\mathbf{1}_{Q_n} \middle| \mathcal{F}_{n-l_n}\right) = (\cosh(t))^{l_n} \times \mathrm{P}\big(Q_n | \mathcal{F}_{n-l_n}\big).$$
Plugging this result into (B.4) gives
$$\mathrm{E}\big(\exp\big(tX_n^{(j)}\big)\mathbf{1}_{Q_n}\big) \le \mathrm{P}(Q_n)(\cosh(t))^{l_n} \mathrm{E}\exp\left(t \max_i X_{n-l_n}^{(i)}\right).$$

Using Lemma 7, we get

$$\mathrm{E}\big(\exp\big(tX_n^{(j)}\big)\mathbf{1}_{Q_n}\big) \le p\mathrm{E}\exp\Big(t\max_i X_n^{(i)}\Big). \tag{B.5}$$

We now study Q_n^c:

$$\begin{aligned}
&\mathrm{E}\bigg(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp\Big(t\max_i X_n^{(i)}\Big)\mathbf{1}_{\{\max_i X_n^{(i)}\ge u_n\}}\mathbf{1}_{Q_n^c}\bigg)\\
&\quad\le p\mathrm{E}\Big(\exp\Big(t\max_i X_n^{(i)}\Big)\mathbf{1}_{Q_n^c}\Big)\\
&\quad\le p\sum_{j=1}^{p}\mathrm{E}\big(\exp\big(tX_n^{(j)}\big)\mathbf{1}_{Q_n^c}\big).
\end{aligned} \tag{B.6}$$

Using the same arguments as before, we get

$$\mathrm{E}\big(\exp\big(tX_n^{(j)}\big)\mathbf{1}_{Q_n^c}\big) \le \mathrm{P}(Q_n^c)(\cosh(t))^{l_n}\mathrm{E}\exp\Big(t\max_i X_{n-l_n}^{(i)}\Big).$$

Using Lemma 7 and the definition of Q_n, we get

$$\begin{aligned}
\mathrm{E}\big(\exp\big(tX_n^{(j)}\big)\mathbf{1}_{Q_n^c}\big) &\le (1-p^{1-p})^{l_n}(\cosh(t))^{l_n}\\
&\qquad\times p(\cosh(t))^{-l_n}\mathrm{E}\exp\Big(t\max_i X_n^{(i)}\Big)\\
&\le p(1-p^{1-p})^{l_n}\mathrm{E}\exp\Big(t\max_i X_n^{(i)}\Big).
\end{aligned} \tag{B.7}$$

Plugging (B.5) into (B.3) and (B.7) into (B.6), together with (B.2), we get

$$\begin{aligned}
&\mathrm{E}\bigg(\frac{d(X_n)}{F_n}\mathbf{1}_{\{X_n\neq 0\}}\exp\Big(t\max_i X_n^{(i)}\Big)\bigg)\\
&\quad\le p\exp(tu_n) + \bigg(\frac{2p^3 l_n}{u_n} + p^3(1-p^{1-p})^{l_n}\bigg)\mathrm{E}\exp\Big(t\max_i X_n^{(i)}\Big).
\end{aligned} \tag{B.8}$$

We note that, owing to Lemma 7,

$$\mathrm{E}\exp\big(tX_n^{(1)}\big) \ge (\cosh(t))^n.$$

Hence, taking $u_n = o(n)$ implies that

$$\exp(tu_n) = o\big(\mathrm{E}\exp\big(tX_n^{(1)}\big)\big).$$

Moreover, using the inequality

$$\mathrm{E}\exp\Big(t\max_i X_n^{(i)}\Big) \le p\mathrm{E}\exp\big(tX_n^{(1)}\big),$$

we see that it is enough to take $l_n = o(u_n)$ and $l_n \to +\infty$ to get the result owing to the estimate (B.8). □

PROOF OF THEOREM 2. By Proposition 4, for all $t \geq 0$, the limit

$$\frac{1}{n} \log \mathrm{E} \exp\left(t X_n^{(1)}\right) \underset{n \to +\infty}{\longrightarrow} \log \cosh(t).$$

Let us consider now a sequence of random variables defined by symmetrization of $X_n^{(1)}$:

$$K_n = \epsilon_n \times X_n^{(1)},$$

where ϵ_n is a symmetric ± 1 Bernoulli random variable independent from $X_n^{(1)}$. Thus, we can recover $X_n^{(1)}$ from K_n owing to the equality $X_n^{(1)} = |K_n|$.

Now, the Laplace transform of K_n is given, for all $t \in \mathbb{R}$, by

$$\begin{aligned} \mathrm{E} \exp(t K_n) &= \tfrac{1}{2} \mathrm{E} \exp\left(t X_n^{(1)}\right) + \tfrac{1}{2} \mathrm{E} \exp\left(-t X_n^{(1)}\right) \\ &= \tfrac{1}{2} \mathrm{E} \exp\left(|t| X_n^{(1)}\right) + \tfrac{1}{2} \mathrm{E} \exp\left(-|t| X_n^{(1)}\right). \end{aligned}$$

When $t = 0$, the above Laplace transform equals 1, whereas, when $t \neq 0$, as n goes to ∞, the term with $+|t|$ dominates (the term with $-|t|$ being bounded from above by 1). As a consequence, using Proposition 4, for *all* $t \in \mathbb{R}$ (not necessarily nonnegative), one has

$$\lim_{n \to +\infty} \frac{1}{n} \log \mathrm{E} \exp(t K_n) = \log \cosh(|t|).$$

The function $\Lambda : t \mapsto \log \cosh(|t|)$ being finite and differentiable everywhere, the Gärtner–Ellis theorem of [6], page 52, entails that the sequence (K_n) satisfies the large-deviations principle with good rate function $\Lambda^*(t)$, where Λ^* is the Legendre transform of Λ. Turning back to $X_n^{(1)} = |K_n|$, the contraction principle (see [6], page 126) implies that $X_n^{(1)}$ satisfies the large-deviations principle with rate function $I = \Lambda^*$ on $[0, 1]$, so Theorem 2 is proved. □

Acknowledgments. We gratefully acknowledge C. Mazza and D. Piau and the anonymous referee for their useful comments.

REFERENCES

[1] BÉRARD, J. and BIENVENÜE, A. (2000). Convergence of a genetic algorithm with finite population. In *Mathematics and Computer Science*: *Algorithms*, *Trees*, *Combinatorics and Probabilities* 155–163. Birkhäuser, Basel.

[2] BÉRARD, J. and BIENVENÜE, A. (2000). Un principe d'invariance pour un algorithme génétique en population finie. *C. R. Acad. Sci. Paris Sér. I Math.* **331** 469–474.

[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[4] BONNAZ, D. (1998). Modèles stochastiques de dynamique des populations. Ph.D. dissertation, Univ. Lausanne.

[5] CERF, R. (1998). Asymptotic convergence of genetic algorithms. *Adv. in Appl. Probab.* **30** 521–550.

[6] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York.

[7] DURRETT, R. (1996). *Probability*: *Theory and Examples*, 2nd ed. Duxbury, Belmont, CA.

[8] GRADSHTEYN, I. S. and RYZHIK, I. M. (1996). *Table of Integrals*, *Series*, *and Products*, 5th ed. Academic, San Diego.

[9] HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Univ. Michigan Press, Ann Arbor.

[10] KESSLER, D., LEVINE, H., RIDGWAY, D. and TSIMRING, L. (1997). Evolution on a smooth landscape. *J. Statist. Phys.* **87** 519–544.

[11] MATSUMOTO, M. and NISHIMURA, T. (1998). Mersenne twister a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Modeling Comput. Simulation* **8** 3–30.

[12] MAZZA, C. and PIAU, D. (2001). On the effect of selection in genetic algorithms. *Random Structures Algorithms* **18** 185–200.

[13] RABINOVICH, Y. and WIGDERSON, A. (1999). Techniques for bounding the convergence rate of genetic algorithms. *Random Structures Algorithms* **14** 111–138.

[14] REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion*. Springer, Berlin.

[15] TSIMRING, L. S., LEVINE, H. and KESSLER, D. A. (1996). RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.* **76** 4440.

LAPCS
UNIVERSITÉ LYON I
BÂTIMENT 101
43 BOULEVARD DU 11 NOVEMBRE 1918
69622 VILLEURBANNE CEDEX
FRANCE
E-MAIL: jean.berard@univ-lyon1.fr

LMC–IMAG
51 RUE DES MATHÉMATIQUES
B.P. 53
38041 GRENOBLE CEDEX 9
FRANCE
E-MAIL: Alexis.Bienvenue@imag.fr