First-Hitting Times for Finite State Spaces

Timo Kötzing and Martin S. Krejca(✉)

Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
martin@krejca.de

Abstract. One of the most important aspects of a randomized algorithm is bounding its expected run time on various problems. Formally speaking, this means bounding the expected first-hitting time of a random process. The two arguably most popular tools to do so are the *fitness level method* and *drift theory*. The fitness level method considers arbitrary transition probabilities but only allows the process to move toward the goal. On the other hand, drift theory allows the process to move into any direction as long as it move closer to the goal in expectation; however, this tendency has to be monotone and, thus, the transition probabilities cannot be arbitrary.

We provide a result that combines the benefit of these two approaches: our result gives a lower and an upper bound for the expected first-hitting time of a random process over $\{0, \dots, n\}$ that is allowed to move forward and backward by 1 and can use arbitrary transition probabilities. In case that the transition probabilities are known, our bounds coincide and yield the exact value of the expected first-hitting time. Further, we also state the stationary distribution as well as the mixing time of a special case of our scenario.

1 Introduction

A very important part of recent research on the theoretical analysis of evolutionary algorithms (EAs) is concerned with run time analysis, and over the years, different tools have been proposed in order to derive run time results more easily. The approaches used for run time analysis all follow the same very broad outline: the algorithm is viewed as a random process whose progress over time is measured. The aim is to bound the expected first-hitting time of the process reaching a certain state, usually finding an optimum. Depending on how much progress can be achieved in different phases of the algorithm, a bound on the expected first-hitting time – the run time – can then be derived. The approaches differ in what phases they consider and how restricted the random process needs to be. We discuss the two arguably most well-known approaches: the *fitness level method* and *drift theory*.

The **fitness level method** is historically older than drift and was first formally defined by Wegener [13] in the context of EAs; a nice overview of this tool including tail bounds was provided by Witt [14]. The method considers a

A. Auger et al. (Eds.): PPSN 2018, LNCS 11102, pp. 79–91, 2018.
https://doi.org/10.1007/978-3-319-99259-4_7

partition of the optimization domain into levels. The progress of an EA is then measured by the expected time it takes to get from one level to another. This means that the expected first-hitting time can be bounded by the sum of the waiting times per level. The major drawback of this tool is that it assumes that there are no cycles among the levels, that is, once the EA advances to a next level, it cannot return to any older level. Hence, this approach basically bounds the first-hitting time of a random walk on a directed path. However, the limitation of having no cycles among the levels results in a very concise theorem that is able to yield *exact* bounds when the actual transition probabilities are known.

Dang and Lehre [1] provide theorems similar to the fitness level method but allow for cycles among the levels. While this approach can yield good upper bounds easily, especially for non-elitist EAs, it assumes that the algorithm of interest makes use of a population. Thus, the theorems cannot be applied to all random processes. Further, without a lower bound, it is not clear how tight a result actually is.

Drift theory is an entirely different approach to deriving expected first-hitting times; see the informative article of Lengler [9] for a general introduction to this topic. Different from the fitness level method, drift theory does not estimate the progress of a random process via waiting times in different levels but instead looks at the expected change of the process after a single step – the *drift*. In this setting, arbitrary steps closer to the goal or away from it may be permitted. The expected first-hitting time then follows from a bound on the drift of the process. Similar to the fitness level method, drift theory can provide upper and lower bounds that are exact if the actual transition probabilities are known.

In its most restrictive setting – the *additive drift theorem* (see Theorem 1), the bound on the drift has to be the same for all states of the random process considered. This means that the bounds on the transition probabilities have to be the same, which limits applicability. In a case where this is overly confining, more advanced theorems like the *variable drift theorem* (see Theorem 2) can be used, which allows to bound the drift dependent on the current state of the process. However, all of these theorems have in common that the the drift needs to be bounded in a monotone way. This means that the drift has to decrease as the goal is approached – a restriction that the fitness level method does not have.

In this paper, we combine the benefits of the fitness level method and of drift theory. Our main result, Theorem 3, considers a random process that is allowed to move toward the goal or away from it in any (not necessarily monotone) fashion. Our setting assumes, in its simplest form, a random walk on an undirected path with the nodes 0 through n. For this setting, we get the exact expected first-hitting time. Our result also provides upper bounds when the process makes larger steps toward the goal and lower bounds when it makes larger steps away from it. We show that our result is a generalization of the fitness level method (Corollary 6) and that it yields bounds that cannot be derived with the variable drift theorem (see Example 4) – the most general drift theorem available. Hence, our result sheds new light on the behavior of random processes over finite state spaces when the progress of the process is not monotone or if the drift is 0.

Further, we also analyze our setting in the context of Markov chains. We give the stationary distribution when the process is a random walk on a path (Theorem 7), again, allowing for arbitrary transition probabilities, and we state an upper bound on the mixing time of the process (Corollary 9). This allows to estimate the probability of the process being in a certain state at a specific point in time – a concept strongly connected to the *any-time analysis* introduced by Jansen and Zarges [6] and similar to *occupation probabilities* as discussed by Lissovoi and Witt [11] and Kötzing et al. [8].

Our paper is structured as follows: Sect. 2 introduces the setting we consider as well as the tools we need in order to derive our results. Section 3 contains our main result, Theorem 3, as well as examples of how the bounds following from it cannot be achieved via any known drift theorem. Last, in Sect. 4, we consider the stationary distribution and the mixing time of the processes we consider.

Note that a special case of Theorem 3 has already been proven by Droste et al. [2] when the Markov chain is a path. Our result extends theirs by providing an upper and a lower bound for scenarios where more transition probabilities are allowed. Further, our result is proven using drift theory, a modern tool that was not available to Droste et al. back then.

2 Setting

We consider random processes $(X_t)_{t\in\mathbb{N}}$ over the finite set $\{0,\dots,n\}$, for an $n \in \mathbb{N}$. In its simplest form, the process is only allowed to move from state s to $s-1$, s, and $s+1$ (if they exist). However, our main result (Theorem 3) generalizes to settings where the process can additionally either make arbitrary long jumps to the front (that is, from state s to any state $s' < s$) or to the back. Our process can be thought of as a random walk as seen in Fig. 1. In the most-restricted scenario, where the process can only move to neighboring states, it performs a random walk on a path.

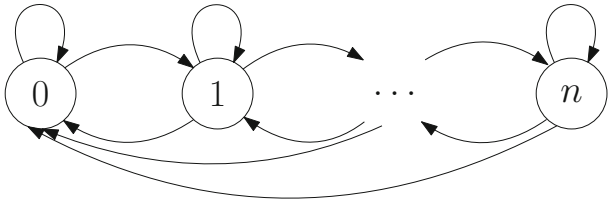

Fig. 1. An exemplary setting we consider. Each node represents a state, and each edge represents a possible transition. Here, the process can move from a state s to any state $s' < s$ when moving to the left but can only move to state $s+1$ when moving to the right.

We are interested in the expected first-hitting time of such a process reaching the state 0. More formally, let $T = \min\{t \mid X_t = 0\}$; we want to bound $\mathrm{E}[T]$.

In our results, we give bounds for $\mathrm{E}[T \mid X_0]$ instead, which is a random variable, as X_0 is a random variable. Bounds on $\mathrm{E}[T]$ can then be derived by the law of total expectation, that is, $\mathrm{E}[T] = \mathrm{E}\big[\mathrm{E}[T \mid X_0]\big]$.

In order to be able to actually reach our goal state 0, we assume that the probability of the random process X_t to move left is positive for any state $s > 0$, that is, $\Pr[X_t - X_{t+1} \geq 1 | X_t = s] > 0$.

2.1 Stochastic Tools

Our results make use of drift theory – a tool that allows to estimate the first-hitting time of a random process when given only estimates of local changes of that process.

The main theorem we use is the following additive drift theorem by He and Yao [3,4]. It yields bounds on the first-hitting time of a random process reaching 0 when the expected local change – the *drift* – can be bounded by a value independent of the current state. We use this theorem in order to prove our main result.

Theorem 1 (Additive Drift [3,5]). *Let $(X_t)_{t\in\mathbb{N}}$ be nonnegative random variables over a finite space $S \subset \mathbb{R}_{\geq 0}$ containing 0, and let $T = \min\{t \mid X_t = 0\}$.*

If there is a constant $\delta > 0$ such that, for all $s \in S$ and all $t < T$,

$$\mathrm{E}[X_t - X_{t+1} \mid X_t = s] \geq \delta, \text{ then } \mathrm{E}[T \mid X_0] \leq \frac{X_0}{\delta}.$$

And if there is a $\delta > 0$ such that, for all $s \in S$ and all $t < T$,

$$\mathrm{E}[X_t - X_{t+1} \mid X_t = s] \leq \delta, \text{ then } \mathrm{E}[T \mid X_0] \geq \frac{X_0}{\delta}.$$

A more flexible drift theorem is the following variable drift theorem. It allows to upper-bound the expected first-hitting time of a random process reaching 0 when the drift can depend in any monotone fashion on the current state. We use this theorem to compare our main result against.

Theorem 2 (Variable Drift [7,12]). *Let $(X_t)_{t\in\mathbb{N}}$ be nonnegative random variables over $\{0\} \cup S$, where $S \subset \mathbb{R}_{\geq 1}$ is a finite state space containing 1, and let $T = \min\{t \mid X_t < 1\}$.*

If there exists a monotonically increasing function $h\colon \mathbb{R}^+ \to \mathbb{R}_{\geq 0}$ such that $1/h$ is integrable and, for all $s \in S$ and all $t < T$,

$$\mathrm{E}[X_t - X_{t+1} \mid X_t = s] \geq h(s), \text{ then } \mathrm{E}[T \mid X_0] \leq \frac{1}{h(1)} + \int_1^{X_0} \frac{1}{h(x)}\,\mathrm{d}x.$$

3 General First-Hitting Times

We start by stating and discussing our main result, Theorem 3, which provides an upper and a lower bound of the first-hitting time of a random process in the setting described in Sect. 2. Those bounds make use of bounds on the transition probabilities of the process.

Theorem 3. *Let $(X_t)_{t\in\mathbb{N}}$ be a random process over $\{0,\dots,n\}$ and let T denote the first point in time t such that $X_t = 0$.*

1. *Suppose there are two functions $p^{\leftarrow}\colon \{1,\dots,n\} \to [0,1]$ and $p^{\rightarrow}\colon \{0,\dots,n-1\} \to [0,1]$ such that, for all $t < T$ and all $s \in \{1,\dots,n\}$,*
   - $p^{\leftarrow}(s) > 0$,
   - $\Pr[X_t - X_{t+1} \geq 1 | X_t = s] \geq p^{\leftarrow}(s)$,
   - $\Pr[X_t - X_{t+1} = -1 | X_t = s] \leq p^{\rightarrow}(s)$ *(for $s \neq n$), and*
   - $\Pr[X_t - X_{t+1} < -1 | X_t = s] = 0$ *(for $s \neq n$).*

   Then
$$\mathrm{E}[T \mid X_0] \leq \sum_{s=1}^{X_0} \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)}.$$

2. *Suppose there are two functions $p^{\leftarrow}\colon \{1,\dots,n\} \to [0,1]$ and $p^{\rightarrow}\colon \{0,\dots,n-1\} \to [0,1]$ such that, for all $t < T$ and all $s \in \{1,\dots,n\}$,*
   - $p^{\leftarrow}(s) > 0$,
   - $\Pr[X_t - X_{t+1} = 1 | X_t = s] \leq p^{\leftarrow}(s)$,
   - $\Pr[X_t - X_{t+1} > 1 | X_t = s] = 0$, *and*
   - $\Pr[X_t - X_{t+1} \leq -1 | X_t = s] \geq p^{\rightarrow}(s)$ *(for $s \neq n$).*

   Then
$$\mathrm{E}[T \mid X_0] \geq \sum_{s=1}^{X_0} \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)}.$$

The bounds on the expected first-hitting time given in Theorem 3 can be thought of as the sum of waiting times. Each waiting time is weighted with the ratio of how likely it is to go away from the goal 0 ($p^{\rightarrow}$) versus going toward it ($p^{\leftarrow}$). Note that the inner sum adds all waiting times up to n. This is where we need that the state space is bounded.

For case 1, note that it does not matter how far left the process moves. In fact, in the proof, we assume the worst case of the process only moving one step closer to the goal. However, we need to guarantee that we can move at most one step away from the goal. The converse is true for case 2: here, we need to guarantee that the process can only move a single step closer to the goal but is allowed to go arbitrarily far away (given its finite state space). Consequently, if the exact transition probabilities are known and, when in state s, the process can only move to the states $s-1$, s, and $s+1$ (if possible), both cases coincide and Theorem 3 yields the exact first-hitting time of the process.

The proof of Theorem 3 is an application of Theorem 1 with a scaled process (a potential) such that the drift can be bounded by 1. The expected first-hitting time is then bounded by the potential of the starting state.

Proof (of Theorem 3). For both cases, we define a potential function $\phi\colon \{0,\dots,n\} \to \mathbb{R}_{\geq 0}$, for $s \in \{0,\dots,n\}$, as follows, using the respective definitions of $p^{\leftarrow}$ and $p^{\rightarrow}$:
$$\phi(s) = \sum_{i=1}^{s} g(i), \text{ where } g(s) = \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} \text{ for } s \neq 0.$$

Note that ϕ is monotonically increasing and that, for all $t < T$, $\phi(X_t) = 0$ if and only if $X_t = 0$. Thus, the first point in time t such that $\phi(X_t) = 0$ is T.

We prove that the following recursion holds via downward induction over $s \in \{1, \dots, n\}$:

$$g(n) = \frac{1}{p^{\leftarrow}(n)} \text{ and } g(s) = \frac{1}{p^{\leftarrow}(s)} + \frac{p^{\rightarrow}(s)}{p^{\leftarrow}(s)} g(s+1) \text{ for } s \neq n.$$

For the base case $s = n$, we get $g(n) = \sum_{i=n}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=n}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} = \frac{1}{p^{\leftarrow}(n)}$, which is true. As for the inductive step, for $s \neq n$, we get

$$\begin{aligned}
&\frac{1}{p^{\leftarrow}(s)} + \frac{p^{\rightarrow}(s)}{p^{\leftarrow}(s)} g(s+1) = \frac{1}{p^{\leftarrow}(s)} + \frac{p^{\rightarrow}(s)}{p^{\leftarrow}(s)} \sum_{i=s+1}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s+1}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} \\
&= \frac{1}{p^{\leftarrow}(s)} + \sum_{i=s+1}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} = \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} = g(s).
\end{aligned}$$

Consider case 1. We first compute the drift for $t < T$ and for $X_t = n$:

$$\begin{aligned}
\mathrm{E}[\phi(X_t) - \phi(X_{t+1}) \,|\, X_t = n] &\geq \Pr X_t - X_{t+1} \geq 1[X_t = n] \cdot \big(\phi(n) - \phi(n-1)\big) \\
&\geq p^{\leftarrow}(n) \cdot \big(\phi(n) - \phi(n-1)\big) = p^{\leftarrow}(n) \cdot g(n) = 1,
\end{aligned}$$

where the first inequality follows from the monotonicity of ϕ.

For $s \in \{1, \dots, n-1\}$ and $t < T$, we get

$$\begin{aligned}
&\mathrm{E}[\phi(X_t) - \phi(X_{t+1}) \,|\, X_t = s] \geq \Pr[X_t - X_{t+1} \geq 1 | X_t = s] \cdot \big(\phi(s) - \phi(s-1)\big) \\
&\qquad + \Pr[X_t - X_{t+1} = -1 | X_t = s] \cdot \big(\phi(s) - \phi(s+1)\big) \\
&= \Pr[X_t - X_{t+1} \geq 1 | X_t = s] \cdot g(s) - \Pr[X_t - X_{t+1} = -1 | X_t = s] \cdot g(s+1) \\
&\geq p^{\leftarrow}(s) \cdot g(s) - p^{\rightarrow}(s) \cdot g(s+1) \\
&= p^{\leftarrow}(s) \cdot \left(\frac{1}{p^{\leftarrow}(s)} + \frac{p^{\rightarrow}(s)}{p^{\leftarrow}(s)} g(s+1) \right) - p^{\rightarrow}(s) \cdot g(s+1) = 1,
\end{aligned}$$

using our recursion scheme for g. Again, the first inequality follows from the monotonicity of ϕ.

Since we have a drift of at least 1 in all cases and a bounded step size, we can apply Theorem 1 and get the desired result.

For case 2, we can perform analogous estimations for the drift but into the other direction, making use that $-\phi(s+1)$ is an upper bound for $-\phi(s')$ for all $s' \geq s+1$. This way, we can upper-bound the drift by 1, yielding the respective lower bound when using Theorem 1. □

Note that the recursion of function g given in the proof is defined as an upward recursion. This actually follows from reconstructing how g has to look in order for the drift to be 1. This approach *cannot* be done in this fashion with a downward recursion, using state 1 as base case, as it is not clear what the

potential for that state has to be, since it has two neighboring states. Thus, it is very important for the search space to be bounded, leading to a well defined base case of $g(n) = 1/p^{\leftarrow}(n)$.

We highlight the importance of the upper bound on the state space (that is, its finiteness) in the following example, where we show that Theorem 2 cannot be easily extended such that it works in a scenario where its drift function h is not monotone.

Example 4. Consider two functions $p^{\leftarrow}\colon \{1,\dots,n\} \to [0,1]$ and $p^{\rightarrow}\colon \{0,\dots,n-1\} \to [0,1]$ such that, for all $s \in \{1,\dots,n\}$,

- $p^{\leftarrow}(s) = \frac{1}{2s}$ (for $s \neq 0$) and
- $p^{\rightarrow}(s) = \frac{1}{2(s+1)}$ (for $s \neq n$).

Note that, for all $s \in \{1,\dots,n-1\}$, $p^{\leftarrow}(s) + p^{\rightarrow}(s) \leq 1$.

Let $(X_t)_{t\in\mathbb{N}}$ be a random process over $\{0,\dots,n\}$ and let T denote the first point in time t such that $X_t = 0$. Suppose that, for all $s \in \{1,\dots,n\}$,

- $\Pr[X_t - X_{t+1} = 1 | X_t = s] = p^{\leftarrow}(s)$ (for $s \neq 0$),
- $\Pr[X_t - X_{t+1} = -1 | X_t = s] = p^{\rightarrow}(s)$ (for $s \neq n$), and
- $\Pr[X_t - X_{t+1} = 0 | X_t = s] = 1 - p^{\leftarrow}(s) - p^{\rightarrow}(s)$.

First, we consider the drift of this process. For all $s \in \{1,\dots,n-1\}$, we get

$$\mathrm{E}[X_t - X_{t+1} \,|\, X_t = s] = \frac{1}{2s} - \frac{1}{2(s+1)} = \frac{s+1}{2s(s+1)} - \frac{s}{2s(s+1)} = \frac{1}{2s(s+1)},$$

and for $s = n$, we get $\mathrm{E}[X_t - X_{t+1} \,|\, X_t = n] = \frac{1}{2n} \geq \frac{1}{2n(n+1)}$. Thus, the drift is dependent on the current state of the process. Note that this dependency is *not* monotonically increasing. However, we ignore this and apply Theorem 2 anyway. Hence, defining $h(s) = 1/\big(2s(s+1)\big)$, we get

$$\begin{aligned}\mathrm{E}[T \,|\, X_0] &\leq \frac{1}{h(1)} + \int_1^{X_0} \frac{1}{h(x)}\mathrm{d}x = 2 \cdot 2 + \int_1^{X_0} 2x(x+1)\mathrm{d}x \\ &= 4 + \frac{2}{3}x^3\Big|_1^{X_0} + x^2|_1^{X_0} = 4 + \frac{2}{3}X_0^3 - \frac{2}{3} + X_0^2 - 1 = \mathrm{O}\big(X_0^3\big).\end{aligned}$$

We now contrast this result with the result following from Theorem 3. Note that our functions $p^{\leftarrow}$ and $p^{\rightarrow}$ are *equal* to the transition probabilities of our random process. Thus, Theorem 3 yields an exact result, as the upper and lower bound coincide:

$$\mathrm{E}[T \,|\, X_0] = \sum_{s=1}^{X_0} \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)}.$$

First, we calculate the product in the expected first-hitting time:

$$\prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} = \prod_{j=s}^{i-1} \frac{\frac{1}{2(j+1)}}{\frac{1}{2j}} = \prod_{j=s}^{i-1} \frac{j}{j+1} = \frac{s}{i},$$

as this is a telescope product and the numerator and denominator of neighboring factors cancel out.

As for the inner sum, we now get $\sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} = \sum_{i=s}^{n} 2i\frac{s}{i} = \sum_{i=s}^{n} 2s = 2s(n-s+1)$, since the sum is independent of its summation index i.

Last, for the outer sum, we get

$$\sum_{s=1}^{X_0} \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} = \sum_{s=1}^{X_0} 2s(n-s+1) = 2\left((n+1)\sum_{s=1}^{X_0} s - \sum_{s=1}^{X_0} s^2\right)$$
$$= 2\left((n+1)\frac{X_0(X_0+1)}{2} - \frac{X_0(X_0+1)(2X_0+1)}{6}\right) = \Theta\big(nX_0^2\big),$$

because $0 \leq X_0 \leq n$, which means that the minuend dominates the difference in the second-to-last line. Overall, Theorem 3 yields $\mathrm{E}[T \mid X_0] = \Theta\big(nX_0^2\big)$.

If we compare this result against the expected first-hitting time due to Theorem 2 of $\mathrm{E}[T \mid X_0] = \mathrm{O}\big(X_0^3\big)$, we see that these results contradict one another if $X_0 = \mathrm{o}(n)$. In fact, if we choose $X_0 = 1$, that is, we are almost at our goal of 0 and have a constant probability of reaching it, the (erroneous) result of Theorem 2 yields a constant first-hitting time, whereas the truth is a first-hitting time linear in n.

Intuitively, this drastic difference comes from the high probability of the process going away from the goal instead of toward it. Thus, if our process does not go toward 0, it may take some time until it returns to 1. Even more important: this waiting time until returning to 1 is dependent on the size of the search space, namely n, as evident by the factor of n in the first-hitting time. Thus, if our search space were unbounded, the expected first-hitting time would be unbounded too, as the probability of returning to 1 would be too small.

This has an even bigger impact on Theorem 2: its result does *not* include the size of the search space.[1] This means that the theorem is inherently not capable of yielding the correct expected first-hitting time in the form given.

When choosing $X_0 = \Theta(n)$, the results of both theorems coincide. In this case, the process starts so far away from the goal 0 that the return time to X_0 is negligible. However, note that this is again due to the search space being bounded. As we start close to the upper bound n, it either takes a short time to return to X_0 (if going away from 0) or we approach the goal. ■

Since Example 4 does not give different results when $X_0 = \Theta(n)$, we provide another example, where the difference in the bounds from Theorems 2 and 3 is tremendous.

Example 5. Consider two functions $p^{\leftarrow}\colon \{1,\dots,n\} \to [0,1]$ and $p^{\rightarrow}\colon \{0,\dots,n-1\} \to [0,1]$ such that, for all $s \in \{1,\dots,n\}$,

- $p^{\leftarrow}(s) = \frac{1}{2}\big(1 + \frac{1}{\mathrm{e}^s}\big)$ (for $s \neq 0$) and
- $p^{\rightarrow}(s) = \frac{1}{2}\big(1 - \frac{1}{\mathrm{e}^s}\big)$ (for $s \neq n$).

[1] The theorem itself assumes the search space to be bounded. However, the actual size of the search space does not matter for the expected first-hitting time.

Note that, for all $s \in \{1, \dots, n-1\}$, $p^{\leftarrow}(s) + p^{\rightarrow}(s) = 1$.

Let $(X_t)_{t \in \mathbb{N}}$ be a random process over $\{0, \dots, n\}$ and let T denote the first point in time t such that $X_t = 0$. Suppose that, for all $s \in \{1, \dots, n\}$,

- $\Pr[X_t - X_{t+1} = 1 | X_t = s] = p^{\leftarrow}(s)$ (for $s \neq 0$),
- $\Pr[X_t - X_{t+1} = -1 | X_t = s] = p^{\rightarrow}(s)$ (for $s \neq n$), and
- $\Pr[X_t - X_{t+1} = 0 | X_t = n] = 1 - p^{\leftarrow}(n)$.

Thus, X_t is almost an unbiased random walk on a path. Further, we assume that $X_0 = n$.

First, we consider the drift of this process. For all $s \in \{1, \dots, n-1\}$, we get $\mathrm{E}[X_t - X_{t+1} \mid X_t = s] = \frac{1}{2}\big(1 + \frac{1}{e^s}\big) - \frac{1}{2}\big(1 - \frac{1}{e^s}\big) = \frac{1}{e^s}$, and for $s = n$, we get $\mathrm{E}[X_t - X_{t+1} \mid X_t = n] = \frac{1}{2}\big(1 + \frac{1}{\mathrm{e}^n}\big) \geq \frac{1}{\mathrm{e}^n}$.

By wrongly applying Theorem 2 with $h(s) = 1/\mathrm{e}^s$, we get $\mathrm{E}[T \mid X_0] \leq \frac{1}{h(1)} + \int_1^{X_0} \frac{1}{h(x)} \mathrm{d}x = \mathrm{e} + \mathrm{e}^n - \mathrm{e} = \mathrm{e}^n$.

We now consider the application of Theorem 3. First, we estimate the product $\prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} \leq 1$. Hence, for the inner sum, we get $\sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} \leq \sum_{i=s}^{n} \frac{2\mathrm{e}^i}{\mathrm{e}^i + 1} \leq \sum_{i=s}^{n} 2 = 2(n - s + 1)$.

Last, for the outer sum, we get

$$\sum_{s=1}^{X_0} \sum_{i=s}^{n} \frac{1}{p^{\leftarrow}(i)} \prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)} \leq \sum_{s=1}^{X_0} 2(n - s + 1) = 2\left((n+1)\sum_{s=1}^{n} 1 - \sum_{s=1}^{n} s\right) = \mathrm{O}\big(n^2\big).$$

Overall, Theorem 3 yields $\mathrm{E}[T \mid X_0] = \mathrm{O}\big(n^2\big)$.

Comparing this with the bound of $\mathrm{O}(\mathrm{e}^n)$ when (wrongly) applying Theorem 2, we see that there is an exponential gap between both results. The result from Theorem 2 is not wrong but nonetheless very much off from the truth. Due to the exponentially declining drift, the bound is exponential. However, the actual first-hitting time of X_t is dominated by the first-hitting time of an unbiased random walk, which hits 0, starting from n, within $\Theta\big(n^2\big)$ steps in expectation. The result from Theorem 3 conforms to this argument. ■

Theorem 3 allows for arbitrary transition probabilities, as long as the state 0 can be reached. If we now restrict the transition probabilities such that the process cannot move to the right, we end up in a scenario where the process can either move closer to the target or stay at its current position. Thus, the expected first-hitting time is the sum of geometrically distributed random variables denoting the number of steps until each state is left. This way, we reconstruct the *fitness level method*.

Corollary 6 (Fitness Level Method [13]). *Let $(X_t)_{t \in \mathbb{N}}$ be a random process over $\{0, \dots, n\}$ and let T denote the first point in time t such that $X_t = 0$. Suppose, for all $t < T$, $X_t - X_{t+1} \geq 0$.*

1. *Suppose there exists a function $p^{\leftarrow} \colon \{1, \dots, n\} \to [0, 1]$ such that, for all $t < T$ and all $s \in \{1, \dots, n\}$,*

- $p^{\leftarrow}(s) > 0$ *and*
- $\Pr[X_t - X_{t+1} \geq 1 | X_t = s] \geq p^{\leftarrow}(s)$.

Then $\mathrm{E}[T \mid X_0] \leq \sum_{s=1}^{X_0} \frac{1}{p^{\leftarrow}(s)}$.

2. *Suppose there exists a function* $p^{\leftarrow} \colon \{1, \dots, n\} \to [0,1]$ *such that, for all* $t < T$ *and all* $s \in \{1, \dots, n\}$,
   - $p^{\leftarrow}(s) > 0$,
   - $\Pr[X_t - X_{t+1} = 1 | X_t = s] \leq p^{\leftarrow}(s)$, *and*
   - $\Pr[X_t - X_{t+1} > 1 | X_t = s] = 0$.

   Then $\mathrm{E}[T \mid X_0] \geq \sum_{s=1}^{X_0} \frac{1}{p^{\leftarrow}(s)}$.

Proof. Both inequalities directly follow from Theorem 3 by noting that the product $\prod_{j=s}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j)}$ is 0 for each $i \geq s+1$, and 1 for $i = s$. □

Note how case 2 assumes that the process can only move one step closer to the goal. If this were not the case, the process could reach the goal 0 earlier (for example, directly from X_0) and we had not to sum over all states between 0 and X_0.

4 Limit Distributions and Mixing Times

Our setting described in Sect. 2 can be interpreted as a Markov chain as depicted in Fig. 1. In this section, we are going to analyze our random process with respect to tools from the theory of Markov chains. We assume that the reader is familiar with the standard terminology in this topic and point to *Markov Chains and Mixing Times* [10] for a nice reference.

In Sect. 3, we determined the expected first-hitting time of a random process on a finite state space. Now we focus on the probability of being in a certain state after a certain time. More specifically, we are interested in a stationary distribution of our process as well as its mixing time. We start with determining a stationary distribution.

In this section, we assume that the process can only move to neighboring states. That is, when in state s, the process can only move to $s-1$, s, and $s+1$ (if possible).

According to Corollary 1.17 from [10], a stationary distribution of a Markov chain is unique if the chain is irreducible, that is, every state can be reached from any other state with positive probability. As we are interested in unique stationary distributions, our following theorem assumes that all transition probabilities to neighboring states are positive.

Theorem 7. *Let* $(X_t)_{t \in \mathbb{N}}$ *be a random process over* $\{0, \dots, n\}$. *Suppose, for all* $t \in \mathbb{N}$, $(X_t - X_{t+1}) \in \{-1, 0, 1\}$. *Suppose there are two functions* $p^{\leftarrow} \colon \{1, \dots, n\} \to [0,1]$ *and* $p^{\rightarrow} \colon \{0, \dots, n-1\} \to [0,1]$ *such that, for all* $s \in \{0, \dots, n\}$,

- $\Pr[X_t - X_{t+1} = 1 | X_t = s] = p^{\leftarrow}(s) > 0$ *(for* $s \neq 0$*) and*
- $\Pr[X_t - X_{t+1} = -1 | X_t = s] = p^{\rightarrow}(s) > 0$ *(for* $s \neq n$*).*

Further, let π denote the stationary distribution of X_t. Then, for all $s \in \{0, \dots, n\}$,

$$\pi(s) = \frac{\prod_{i=0}^{s-1} \frac{p^{\rightarrow}(i)}{p^{\leftarrow}(i+1)}}{\sum_{i=0}^{n} \prod_{j=0}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j+1)}}.$$

Similar to Theorem 3, the ratio of the transition probabilities are very important for our result. However, different from Theorem 3, we now need the values $p^{\rightarrow}(s)/p^{\leftarrow}(s+1)$ instead of $p^{\rightarrow}(s)/p^{\leftarrow}(s)$. This difference makes sense: in Sect. 3, we were interested in reaching the state 0. Thus, it was important how likely the process moves toward or from the goal. Now, there is no special state that we want to reach. We are interested in the probability of being in a certain state. Hence, it is important with which probability to get to a state and with which probability to leave it again.

If the ratio $p^{\rightarrow}(s)/p^{\leftarrow}(s+1)$ is the same for all states s, we can simplify the stationary distribution as follows.

Corollary 8. *Let $(X_t)_{t \in \mathbb{N}}$ be a random process over $\{0, \dots, n\}$. Suppose, for all $t \in \mathbb{N}$, $(X_t - X_{t+1}) \in \{-1, 0, 1\}$. Further, suppose there are two functions $p^{\leftarrow} \colon \{1, \dots, n\} \to [0, 1]$ and $p^{\rightarrow} \colon \{0, \dots, n-1\} \to [0, 1]$ and a value $c > 0$ such that, for all $s \in \{0, \dots, n\}$,*

- $\Pr[X_t - X_{t+1} = 1 | X_t = s] = p^{\leftarrow}(s) > 0$ *(for $s \neq 0$),*
- $\Pr[X_t - X_{t+1} = -1 | X_t = s] = p^{\rightarrow}(s) > 0$ *(for $s \neq n$), and*
- $p^{\rightarrow}(s)/p^{\leftarrow}(s+1) = c$.

Further, let π denote the stationary distribution of X_t. If $c \neq 1$, then, for all $s \in \{0, \dots, n\}$, it holds that $\pi(s) = (c-1)\frac{c^s}{c^{n+1}-c}$.

And if $c = 1$, for all $s \in \{0, \dots, n\}$, it holds that $\pi(s) = \frac{1}{n+1}$.

Proof. We use the definition of π from Theorem 7 and use that $\prod_{j=0}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j+1)} = \prod_{j=0}^{i-1} c = c^i$. Hence, we get $\pi(s) = \frac{\prod_{i=0}^{s-1} \frac{p^{\rightarrow}(i)}{p^{\leftarrow}(i+1)}}{\sum_{i=0}^{n} \prod_{j=0}^{i-1} \frac{p^{\rightarrow}(j)}{p^{\leftarrow}(j+1)}} = \frac{c^s}{\sum_{i=0}^{n} c^i}$, where the result follows by noting that the denominator is a geometric sum (when $c \neq 1$). For $c = 1$, the result is trivial. □

When we have $c < 1$, that is, the probability to move to the right is less than the probability to move to the left, Corollary 8 yields that the probability to be in state s declines exponentially in s. Conversely, if $c > 1$, the probability grows exponentially in s. Last, if $c = 1$, we end up with the uniform distribution.

Given the stationary distribution, it is a natural question to ask whether this distribution will be reached in the limit of the number of steps going to infinity. This is the case if the Markov chain is also aperiodic (besides irreducible; Theorem 4.9 from [10]), which is the case if there is at least one state that has a self loop (that is, it has a positive probability of reaching itself in one step).

Assuming that our Markov chain is also aperiodic, we now determine its mixing time, that is, the time until the probability to be in state s is only at most ε different from the probability stated by the stationary distribution.

The following corollary is a direct consequence of Corollary 14.7 from [10]. The idea behind the corollary is to consider a coupling of two independent copies of the process that start maximally far apart. The goal is that both processes meet. If they arrive in neighboring states and at least one of those states has a self loop, this is possible by one process staying where it is and the other process moving to said state. In order for this argument to translate into a mixing time, it is necessary that the expected distance of two such neighboring processes is less than 1 after one step. That is, in expectation, they move close to one another once they are next to each other. This is formalized in the following corollary.

Corollary 9. *Let $(X_t)_{t\in\mathbb{N}}$ be a random process over $\{0,\dots,n\}$. Suppose, for all $t \in \mathbb{N}$, $(X_t - X_{t+1}) \in \{-1,0,1\}$. Further, suppose there are two functions $p^{\leftarrow}\colon \{1,\dots,n\} \to [0,1]$ and $p^{\rightarrow}\colon \{0,\dots,n-1\} \to [0,1]$ such that, for all $s \in \{0,\dots,n\}$,*

- $\Pr[X_t - X_{t+1} = 1 | X_t = s] = p^{\leftarrow}(s) > 0$ *(for $s \neq 0$) and*
- $\Pr[X_t - X_{t+1} = -1 | X_t = s] = p^{\rightarrow}(s) > 0$ *(for $s \neq n$).*

Let $(Y_t)_{t\in\mathbb{N}}$ be an independent copy of X_t. Assume that there is an $\alpha > 0$ such that, for all $s \in \{0,\dots,n-1\}$,

$$\mathrm{E}\big[|X_{t+1} - Y_{t+1}| \,\big|\, X_t = s, Y_t = s+1\big] \leq \mathrm{e}^{-\alpha}. \tag{1}$$

Let t_{mix} denote the mixing time of X_t. Then, for each $\varepsilon \in (0,1/2)$,

$$t_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{\ln \frac{n}{\varepsilon}}{\alpha} \right\rceil.$$

Proof. The statement is an application of Corollary 14.7 from [10]. Inequality (1) is a special case of the situation considered in Theorem 14.6 in [10]. □

Note that if we consider a state s such that both s and $s+1$ have no self loop, $\mathrm{E}\big[|X_{t+1} - Y_{t+1}| \,\big|\, X = s, Y_t = s+1\big]$ is at least 1, as X_{t+1} and Y_{t+1} cannot be in the same state. Thus, Corollary 9 is not applicable in this scenario.

References

1. Dang, D.C., Lehre, P.K.: Runtime analysis of non-elitist populations: from classical optimisation to partial information. Algorithmica **75**(3), 428–461 (2016)
2. Droste, S., Jansen, T., Wegener, I.: Dynamic parameter control in simple evolutionary algorithms. In: Proceedings of the FOGA 2000, pp. 275–294 (2000)
3. He, J., Yao, X.: Drift analysis and average time complexity of evolutionary algorithms. Artif. Intell. **127**(1), 57–85 (2001)
4. He, J., Yao, X.: Erratum to: drift analysis and average time complexity of evolutionary algorithms. Artif. Intell. **140**(1), 245–248 (2002)

5. He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Nat. Comput. **3**(1), 21–35 (2004)
6. Jansen, T., Zarges, C.: Evolutionary algorithms and artificial immune systems on a bi-stable dynamic optimisation problem. In: Proceedings of the GECCO 2014, pp. 975–982 (2014)
7. Johannsen, D.: Random combinatorial structures and randomized search heuristics. Ph.D. thesis, Universität des Saarlandes (2010). http://scidok.sulb.uni-saarland.de/volltexte/2011/3529/pdf/Dissertation3166JohaDani2010.pdf
8. Kötzing, T., Lissovoi, A., Witt, C.: (1 + 1) EA on generalized dynamic OneMax. In: Proceedings of the FOGA XIII, pp. 40–51 (2015)
9. Lengler, J.: Drift analysis. CoRR abs/1712.00964 (2017). http://arxiv.org/abs/1712.00964
10. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society, Providence (2006)
11. Lissovoi, A., Witt, C.: MMAS versus population-based EA on a family of dynamic fitness functions. Algorithmica **75**(3), 554–576 (2016)
12. Mitavskiy, B., Rowe, J.E., Cannings, C.: Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links. Int. J. Intell. Comput. Cybern. **2**(2), 243–284 (2009)
13. Wegener, I.: Theoretical aspects of evolutionary algorithms. In: Orejas, F., Spirakis, P.G., van Leeuwen, J. (eds.) ICALP 2001. LNCS, vol. 2076, pp. 64–78. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-48224-5_6
14. Witt, C.: Fitness levels with tail bounds for the analysis of randomized search heuristics. Inf. Process. Lett. **114**(1–2), 38–41 (2014)