Theoretical Computer Science 796 (2019) 51–69

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

First-hitting times under drift

Timo Kötzing, Martin S. Krejca *

Hasso Platter Institute, University of Potsdam, Potsdam, Germany

ARTICLE INFO

Article history:
Received 2 October 2018
Received in revised form 14 May 2019
Accepted 12 August 2019
Available online 20 August 2019
Communicated by L.M. Kirousis

Keywords:
First-hitting time
Random process
Drift

ABSTRACT

For the last ten years, almost every theoretical result concerning the expected run time of a randomized search heuristic used *drift theory*, making it the arguably most important tool in this domain. Its success is due to its ease of use and its powerful result: drift theory allows the user to derive bounds on the expected first-hitting time of a random process by bounding expected local changes of the process – the *drift*. This is usually far easier than bounding the expected first-hitting time directly.

Due to the widespread use of drift theory, it is of utmost importance to have the best drift theorems possible. We improve the fundamental additive, multiplicative, and variable drift theorems by stating them in a form as general as possible and providing examples of why the restrictions we keep are still necessary. Our additive drift theorem for upper bounds only requires the process to be lower-bounded, that is, we remove unnecessary restrictions like a finite, discrete, or bounded state space. As corollaries, the same is true for our upper bounds in the case of variable and multiplicative drift. By bounding the step size of the process, we derive new lower-bounding multiplicative and variable drift theorems. Last, we also state theorems that are applicable when the process has a drift of 0, by using a drift on the variance of the process.

1. Drift theory

In the theory of randomized algorithms, the first and most important part of algorithm analysis is to compute the expected run time. A finite run time guarantees that the algorithm terminates almost surely, and, due to Markov's inequality, the probability of the run time being far larger than the expected value can be bounded, too. Thus, it is important to have strong and easy to handle tools in order to derive expected run times. The de facto standard for this purpose in the theory of randomized search heuristics is *drift theory*.

Drift theory is a general term for a collection of theorems that consider random processes and bound the expected time it takes the process to reach a certain value – the *first-hitting time*. The beauty and appeal of these theorems lie in them usually having few restrictions but yielding strong results. Intuitively speaking, in order to use a drift theorem, one only needs to estimate the expected change of a random process – the *drift* – at any given point in time. Hence, a drift theorem turns expected local changes of a process into expected first-hitting times. In other words, local information of the process is transformed into global information.

Drift theory gained traction in the theory of randomized search heuristics when it was introduced to the community by He and Yao [1,2] via the *additive drift theorem*. However, they were not the first to prove it. The result dates back to

* Corresponding author.
E-mail addresses: timo.koetzing@hpi.de (T. Kötzing), martin.krejca@hpi.de (M.S. Krejca).

https://doi.org/10.1016/j.tcs.2019.08.021

Hajek [3], who stated the theorem in a fashion quite different from how it is phrased nowadays. According to Lengler [4], the theorem has been proven even prior to that various times. Since then, many different versions of drift theorems have been proven, the most common ones being the *variable drift theorem* [5] and the *multiplicative drift theorem* [6]. The different names refer to how the drift is bounded other than independent of time: additive means that the drift is bounded by the same value for all states; in a multiplicative scenario, the drift is bounded by a multiple of the current state of the process; and in the setting of variable drift, the drift is bounded by any monotone function with respect to the current state of the process.

At first, the theorems were only stated over finite or discrete state spaces. However, these restrictions are seldom used in the proofs and thus not necessary, as pointed out, for example, by Lehre and Witt [7], who prove a general drift theorem without these restrictions. Nonetheless, up to date, almost all drift theorems require a bounded state space. Semenov and Terkel [8] state a theorem very much like an additive drift theorem for unbounded state spaces, but they require the process to have a bounded variance, as they also prove concentration for their result. Corus et al. [9] provide a proof of an additive drift theorem over an unbounded[1] state space in their appendix. However, they require that the expected first-hitting time is finite, which is not always easy to prove and thus restricts the applicability of the theorem. Lehre and Witt [10] provide drift theorems for unbounded state spaces when interested in an upper bound on the expected first-hitting time. However, for a lower bound, the state space must be still bounded. Last, Lengler and Steger [11] prove drift theorems for unbounded state spaces but only do so for discrete processes. Our main result (Theorem 7) removes the restriction of a bounded state space and even allows the process to overshoot the target value. In return, the theorem is easy to use yet more permissive than all prior theorems.

We improve the state-of-the-art of drift theory by proving drift theorems over unbounded state spaces. Each theorem comes in two variants: one variant providing an upper bound on the expected first-hitting time, the other variant providing a lower bound. All of our results for upper bounds require the random process to be bounded from below, and for most of them we assume the process to be nonnegative in order to get conciser bounds. In contrast to that, for our results on lower bounds, we allow the process to be completely unbounded, but we have to bound the step size in return; Example 9 shows why our theorems fail otherwise.

Our most important results are our upper and lower bound of the classical additive drift theorem (Theorem 7 and 8, respectively), which we prove for unbounded state spaces. These theorems are used as a foundation for all of our other drift theorems in other settings. In the case of variable and multiplicative drift, we consider two different first-hitting times. Overall, our results can be summarized as follows:

Additive drift: We prove an upper bound for any process (Theorem 7) bounded from below, and a lower bound for processes with bounded *expected* step size (Theorem 8).

Multiplicative drift: We prove upper bounds for any nonnegative process (Corollary 16 and 17), and four different lower bounds for processes with bounded step size (Corollary 19, Corollary 20, Corollary 21, Theorem 23), where we consider different regimes of the starting value of the process.

Variable drift: We prove upper bounds for any nonnegative process (Theorem 10, Theorem 11), and lower bounds for processes with bounded step size and a restriction to the growth of the drift (Theorem 12, Corollary 13, Theorem 14, Corollary 15).

No drift: We provide a theorem (Theorem 25) that transforms a random process with no drift into one with a positive drift. This transformation can be applied to many drift theorems; however, the requirements may be harder to check. We give some examples of how this transformation can be applied in Corollary 26, Corollary 27, and Corollary 28.

The intention of this paper is to provide a fully-packed reference for very general yet easy-to-apply drift theorems. That is, we try to keep the requirements of the theorems as easy as possible but still state the theorems in the most general way, given the restrictions. Further, we discuss the ideas behind the different theorems and some of the proofs in order to provide insights into how and why drift works, we provide examples, and we discuss prior work at the beginning of each section.

We only consider bounds of the expected first-hitting time, as this is already a vast field to explore. However, we want to mention that drift theory has also brought forth other results than expected first-hitting times, namely, concentration bounds and negative drift, which are related. Both areas bound *the probability* of the first-hitting time taking certain values. Concentration bounds show how unlikely it is for a process to take much longer than the expected first-hitting time [10,12,13]. On the other hand, negative drift bounds how likely it is for the process to reach the goal although the drift is going the opposite direction [13,14]. These results are also very helpful but out of the scope of this paper.

Our paper is structured as follows: in Section 2, we start by introducing important notation and terms, which we use throughout the entire paper. Further, we also discuss Theorem 1, which most of our proofs of the additive drift theorems rely on. In Section 3, we discuss additive drift and prove our main results. We then continue with variable drift in Section 4 as a generalization of additive drift. In this section, we introduce two different versions of fist-hitting time that our results are based on. In Section 5, we consider the scenario of multiplicative drift, where, in addition to stating our results, we discuss the tightness of our different lower bounds. Last, in Section 6, we consider processes with a drift of 0.

[1] They still require a lower bound for the state space but not an upper bound. We still refer to such a setting as unbounded.

This paper extends our previous results [15] by extending Theorem 4 to also work for processes with negative values, by proving various lower bounds for more scenarios than additive drift, and by also considering the case when there is no drift.

2. Preliminaries

We consider the expected *first-hitting time* T of a process $(X_t)_{t\in\mathbb{N}}$ over $\mathbb{R}$, which we call X_t for short. That is, we are interested in the expected time it takes the process to reach a certain value for the first time, which we will refer to as the *target*. Usually, our target is the value 0, that is, we will define the random variable $T = \inf\{t \mid X_t \leq 0\}$.

We provide bounds on $\mathrm{E}[T \mid X_0]$ with respect to the *drift* of X_t, which is defined as

$$X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] .$$

Note that $\mathrm{E}[T \mid X_0]$ as well as $\mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t]$ are both random variables. Because of the latter, the drift is a random variable, too. Further note that, if the drift is positive, X_t decreases its value in expectation over time when considering positive starting values. This is why 0 will be our target most of the time.

We are only interested in the process X_t until the time point T. That is, all of our requirements only need to hold for all $t < T$ (since we also consider $t+1$). While this phrasing is intuitive, it is formally inaccurate, as T is a random variable. We will continue to use it, however, formally, each of our inequalities in each of our requirements should be multiplied with the characteristic function of the event $\{t < T\}$. This way, the inequalities trivially hold once $t \geq T$ and, otherwise, are the inequalities we state. This is similar to conditioning on the event $\{t < T\}$ but has the benefit of being valid even if $\Pr[t < T] = 0$ holds.

We want to mention that all of our results actually hold for a random process $(X_t)_{t\in\mathbb{N}}$ adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$, where T is a stopping time defined with respect to $\mathcal{F}_t$. Since this detail is frequently ignored in drift theory, we phrase all of our results with respect to the natural filtration, making them look more familiar to usual drift results. For any time point $t \leq T$, we call $X_0, \ldots, X_{t-1}$ the *history* of the process.

Last, we state all of our results conditional on X_0, that is, we bound $\mathrm{E}[T \mid X_0]$. However, by the law of total expectation, one can easily derive a bound for $\mathrm{E}[T] = \mathrm{E}\big[\mathrm{E}[T \mid X_0]\big]$.

2.1. Martingale theorems

In this section, we state the theorems that we will use in order to prove our results in the next sections. These theorems make use of *martingales*, a fundamental concept in the field of probability theory. A martingale is a random process with a drift of 0, that is, in expectation, it does not change over time. Further, a *supermartingale* has a drift of at least 0, that is, it decreases over time in expectation, and a *submartingale* has a drift of at most 0, that is, it increases over time in expectation.

The arguably most important theorem for martingales is the Optional Stopping Theorem (Theorem 1). It is often only provided in a form that suits martingales. We use a version given by Grimmett and Stirzaker [16, Chapter 12.5, Thm. 9] but extend its use to super- and submartingales, which is possible as mentioned by Bhattacharya and Waymire [17, Remark 3.7].

Theorem 1 *(Optional stopping). Let $(X_t)_{t\in\mathbb{N}}$ be a sequence of random variables over $\mathbb{R}$, and let T be a stopping time[2] for X_t. Suppose that*

(a) $\mathrm{E}[T] < \infty$ *and that*
(b) there is some value $c \geq 0$ such that, for all $t < T$, it holds that $\mathrm{E}\big[|X_{t+1} - X_t| \,\big|\, X_0, \ldots, X_t\big] \leq c$.

Then:

1. *If, for all $t < T$, we have $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq 0$, then*

$$\mathrm{E}[X_T] \leq \mathrm{E}[X_0]$$

2. *If, for all $t < T$, we have $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \leq 0$, then*

$$\mathrm{E}[X_T] \geq \mathrm{E}[X_0] .$$

Theorem 1 allows us to bound $\mathrm{E}[X_T]$ independently of its history, which is why our drift results are independent of the history of X_T as well.

[2] Intuitively, for the natural filtration, a stopping time T is a random variable over $\mathbb{N}$ such that, for all $t \in \mathbb{N}$, the event $\{t \leq T\}$ is only dependent on $X_0, \ldots, X_t$.

Note that case (1) refers to supermartingales, whereas case (2) refers to submartingales. Intuitively, case (1) says that a supermartingale will have, in expectation, a lower value than it started with, which makes sense, as a supermartingale decreases over time in expectation. Case (2) is analogous for submartingales. For martingales, both cases can be combined in order to yield an equality.

Martingales are essential in the proofs of our theorems. We will frequently transform our process such that it results in a supermartingale or a submartingale in order to apply Theorem 1.

For the special case of nonnegative supermartingales, condition (a) of Theorem 1 is not necessary, which results in the following very useful Optional Stopping Theorem.

Theorem 2 *(Optional stopping for nonnegative supermartingales [18, Theorem 4.8.4]). Let $(X_t)_{t\in\mathbb{N}}$ be a sequence of random variables over $\mathbb{R}$, and let T be a stopping time for X_t. Suppose that,*

(a) for all $t \leq T$, we have $X_t \geq 0$ and that,
(b) for all $t \leq T$, we have $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq 0$.

Then

$$\mathrm{E}[X_T] \leq \mathrm{E}[X_0] .$$

Another useful theorem for martingales is the following Azuma–Hoeffding Inequality [19]. This inequality basically is for martingales what a Chernoff bound is for binomial distributions.

Theorem 3 *(Azuma–Hoeffding inequality). Let $(X_t)_{t\in\mathbb{N}}$ be a sequence of random variables over $\mathbb{R}$. Suppose that*

- *there is some value $c > 0$ such that, for all $t \in \mathbb{N}$, we have $|X_t - X_{t+1}| < c$.*

If, for all $t \in \mathbb{N}$, $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq 0$, then, for all $t \in \mathbb{N}$ and all $r > 0$, it holds that

$$\Pr[X_t - X_0 \geq r] \leq \mathrm{e}^{-\frac{r^2}{2tc^2}} .$$

3. Additive drift

We speak of additive drift when the drift can be bounded by a value independent of the process itself. That is, the bound is spatially and time-homogeneous.

When considering the first-hitting time T of a random process $(X_t)_{t\in\mathbb{N}}$ whose drift is *lower*-bounded by a value $\delta > 0$, then $\mathrm{E}[T \mid X_0]$ is *upper*-bounded by X_0/δ. Interestingly, if the drift of X_t is *upper*-bounded by δ, $\mathrm{E}[T \mid X_0]$ is *lower*-bounded by X_0/δ. Thus, if the drift of X_t is exactly δ, that is, we know how much expected progress X_t makes in each step, our expected first-hitting time is equal to X_0/δ. This result is remarkable, as it can be understood intuitively as follows: since we stop once X_t reaches 0, the distance from our start (X_0) to our goal (0) is exactly X_0, and we make an expected progress of δ each step. Thus, in expectation, we are done after X_0/δ steps.

3.1. Upper bounds

We give a proof for the Additive Drift Theorem, originally published (in a more restricted version) by He and Yao [1,2]. We start by reproving the original theorem (which requires a bounded state space) but in a simpler, more elegant and educational manner. We then greatly extend this result by generalizing it to processes with a bounded step width. Finally, we lift also this restriction.

In most of these cases, we assume that our random process is nonnegative and has to hit 0 exactly, as this makes the statements of our theorems conciser. The intuitive reason for this is the following: when estimating an upper bound for the expected first-hitting time, we need a lower bound of the drift. This means the larger our bound of the drift, the better our bound for the first-hitting time. Since our process is nonnegative, the drift for values close to 0 provides a natural bound for the drift (which is uniform over the entire state space, since we look at *additive* drift). If our process could take values less than 0, we could artificially increase our lower bound of the drift for values that are now bounded by 0 and, thus, improve our first-hitting time. Close to the end of this section, we also give an example (Example 6), which shows how the concise statements of our drift theorems fail if the process can take negative values. However, our most general version is capable of handling such cases by incorporating an extra term in the result that compensates for jumps below 0.

The proof of the following theorem transforms the process into a supermartingale and then uses Theorem 1. However, in order to apply Theorem 1, we have to make sure to fulfill condition (a), which is the hardest part.

Theorem 4 *(Upper additive drift, bounded). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, and let $T = \inf\{t \mid X_t \leq 0\}$. Furthermore, suppose that,*

(a) for all $t \leq T$, it holds that $X_t \geq 0$, that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq \delta$, and that
(c) there is some value $c \geq 0$ such that, for all $t < T$, it holds that $X_t \leq c$.

Then

$$\mathrm{E}[T \mid X_0] \leq \frac{X_0}{\delta}.$$

Note that condition (a) means that T can be rewritten as $\inf\{t \mid X_t = 0\}$, that is, we have to hit 0 exactly in order to stop. We show in Example 6 why this condition is crucial.

Condition (b) bounds the expected progress we make each time step. The larger δ, the lower the expected first-hitting time. However, due to condition (a), note that small values of X_t create a natural upper bound for δ, as the progress for such values can be at most $|X_t - 0| = X_t$.

Condition (c) means that we are considering random variables over the interval $[0, c]$. It is a restriction that all previous additive drift theorems have but that is actually not necessary, as we show with Theorem 7. In the following proof, we use this condition in order to show that $\mathrm{E}[T] < \infty$, which is necessary when applying Theorem 1.

Proof of Theorem 4. We want to use case (1) of the Optional Stopping Theorem in the version of Theorem 1. Thus, we define, for all $t < T$, $Y_t = X_t + \delta t$, which is a supermartingale, since

$$\begin{aligned} Y_t - \mathrm{E}[Y_{t+1} \mid Y_0, \ldots, Y_t] &= X_t + \delta t - \mathrm{E}[X_{t+1} + \delta(t+1) \mid X_0, \ldots, X_t] \\ &= X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] - \delta \geq 0, \end{aligned}$$

as we assume that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq \delta$ for all $t < T$. Note that we can change the condition $Y_0, \ldots, Y_t$ to $X_0, \ldots, X_t$ because the transformation from X_t to Y_t is injective.

We now show that $\mathrm{E}[T \mid X_0] < \infty$ holds in order to apply Theorem 1. Let $r > 0$, and let a be any value such that $\Pr[X_0 \leq a] > 0$. We condition on the event $\{X_0 \leq a\}$, and we consider a time point $t' = (a + r)/\delta$ and want to bound the probability that $X_{t'}$ has not reached 0 yet, i.e., the event $\{X_{t'} > 0\}$. We rewrite this event as $\{X_{t'} - a > -a\}$, which is equivalent to $\{Y_{t'} - a > -a + \delta t' = r\}$, by definition of Y and t'.

Note that, for all $t < T$, $|Y_t - Y_{t+1}| < c + \delta + 1$, as we assume that $X_t \leq c$. Thus, the differences of Y_t are bounded and we can apply Theorem 3 as follows, noting that $Y_0 = X_0 \leq a$, due to our condition on $\{X_0 \leq a\}$:

$$\Pr[Y_{t'} - a > r \mid X_0 \leq a] \leq \Pr[Y_{t'} - Y_0 \geq r \mid X_0 \leq a] \leq \mathrm{e}^{-\frac{r^2}{2t'(c+\delta+1)^2}}.$$

If we choose $r \geq a$, we get $t' \leq 2r/\delta$ and, thus,

$$\Pr[Y_{t'} - Y_0 > r \mid X_0 \leq a] \leq \mathrm{e}^{-\frac{r\delta}{4(c+\delta+1)^2}}.$$

This means that the probability that $X_{t'}$ has not reached 0 goes exponentially fast toward 0 as t' (and, hence, r) goes toward ∞. Thus, the expected value of T is finite.

Now we can use case (1) of Theorem 1 in order to get $\mathrm{E}[Y_T \mid X_0] \leq \mathrm{E}[Y_0 \mid X_0]$. In particular, noting that $X_T = 0$ by definition,

$$\begin{aligned} X_0 &= \mathrm{E}[X_0 \mid X_0] = \mathrm{E}[Y_0 \mid X_0] \geq \mathrm{E}[Y_T \mid X_0] = \mathrm{E}[X_T + \delta T \mid X_0] \\ &= \mathrm{E}[X_T \mid X_0] + \delta \mathrm{E}[T \mid X_0] = \delta \mathrm{E}[T \mid X_0]. \end{aligned}$$

Thus, we get the desired bound by dividing by δ. □

Note that the arguments in this proof only need the property of bounded differences in order to apply Theorem 3. Thus, we can relax the condition of a bounded state space into bounded step size, which can be seen in the following corollary.

Corollary 5 *(Upper additive drift, bounded step size). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, and let $T = \inf\{t \mid X_t \leq 0\}$. Furthermore, suppose that,*

(a) for all $t \leq T$, it holds that $X_t \geq 0$, that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq \delta$, and that
(c) there is some value $c \geq 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \leq c$.

Then

$$\mathrm{E}[T \mid X_0] \leq \frac{X_0}{\delta}.$$

As we already mentioned before, note that the condition of the process not being negative is important in order to get correct results. The following example highlights this fact.

Example 6. Let $n > 0$, and let $(X_t)_{t\in\mathbb{N}}$ be a random process with $X_0 = 1$ and, for all $t \in \mathbb{N}$, $X_{t+1} = X_t$ with probability $1 - 1/n$, and $X_{t+1} = -n + 1$ otherwise. Let T denote the first point in time t such that the event $X_t \leq 0$ occurs. We have, for all $t < T$, that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \dots, X_t] = 1$ and, thus, $\mathrm{E}[T \mid X_0] \leq 1$ if we could apply any of the additive drift theorems. However, since T follows a geometric distribution with success probability $1/n$, we have $\mathrm{E}[T \mid X_0] = n$.

The reason that the drift theorems so far fail for processes that also take negative values is that we explicitly used $\mathrm{E}[X_T \mid X_0] = 0$ in the proof. This assumption can, of course, be violated if the process allows negative values. In the following theorem, which is the most general additive drift theorem up to date, we only require some lower bound α on the state space (which may be negative) and phrase the result such that the value of $\mathrm{E}[X_T \mid X_0]$ is incorporated. Further, we use Theorem 2 in order to not have to prove that the process converges almost surely.

It is important to note that we define the first-hitting time T to still be the first point in time when the process reaches 0 or a value below it. The variable α only denotes the lower bound of the state space of the process. This allows for a more general application of the theorem, since 0 does not have to be hit exactly but can also be surpassed. However, this results in $\mathrm{E}[X_T \mid X_0]$ being part of the upper bound for $\mathrm{E}[T \mid X_0]$, which may be hard to bound. Thus, we also state a second bound, which is more coarse but does not involve any extra calculations.

Theorem 7 *(Upper additive drift, unbounded). Let $\alpha \leq 0$, let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, and let $T = \inf\{t \mid X_t \leq 0\}$. Furthermore, suppose that,*

(a) for all $t \leq T$, it holds that $X_t \geq \alpha$, and that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \dots, X_t] \geq \delta$.

Then

$$\mathrm{E}[T \mid X_0] \leq \frac{X_0 - \mathrm{E}[X_T \mid X_0]}{\delta} \leq \frac{X_0 - \alpha}{\delta} .$$

Proof. We use the same proof strategy as for Theorem 4 with the difference being that we are going to apply Theorem 2 instead of Theorem 1. Since Theorem 2 still requires a nonnegative process, we define, for all $t < T$, $Y_t = X_t + \delta t - \alpha$, which is nonnegative for all $t \leq T$, since $Y_t \geq X_t - \alpha \geq 0$, due to condition (a). Further, Y_t is a supermartingale for all $t \leq T$, since

$$\begin{aligned} Y_t - \mathrm{E}[Y_{t+1} \mid Y_0, \dots, Y_t] &= X_t + \delta t - \alpha - \mathrm{E}[X_{t+1} + \delta(t+1) - \alpha \mid X_0, \dots, X_t] \\ &= X_t - \mathrm{E}[X_{t+1} \mid X_0, \dots, X_t] - \delta \geq 0 , \end{aligned}$$

due to condition (b). Hence, we can apply Theorem 2 and get $\mathrm{E}[Y_T \mid X_0] \leq \mathrm{E}[Y_0 \mid X_0]$. Using the definition of Y_t, especially that $\mathrm{E}[X_0 - \alpha \mid X_0] = \mathrm{E}[Y_0 \mid X_0]$, we get

$$\begin{aligned} X_0 - \alpha &= \mathrm{E}[X_0 - \alpha \mid X_0] = \mathrm{E}[Y_0 \mid X_0] \geq \mathrm{E}[Y_T \mid X_0] = \mathrm{E}[X_T + \delta T - \alpha \mid X_0] \\ &= \mathrm{E}[X_T \mid X_0] + \delta \mathrm{E}[T \mid X_0] - \alpha . \end{aligned}$$

Solving this inequality for $\mathrm{E}[T \mid X_0]$ yields the first of the two bounds of this theorem.

For the second bound, we use the bound we just derived and trivially bound $\mathrm{E}[X_T \mid X_0] \geq \alpha$, due to condition (a), which concludes the proof. □

Note that Example 6 is not a counterexample to Theorem 7, as $X_T = -n + 1$, which is incorporated into the bound of the theorem. Applying Theorem 7 states the correct value of $\mathrm{E}[T \mid X_0]$, being n.

We would like to state that we do not know of any counterexample for Theorem 7 when assuming that condition (a) does not hold, that is, when the state space is truly unbounded. We conjecture that the statement of Theorem 7 still holds in this case.

3.2. Lower bounds

In this section, we provide a lower bound for the expected first-hitting time under additive drift. In order to do so, we need an upper bound for the drift. Since we now lower-bound the first-hitting time, a large upper bound of the drift makes the result bad. Thus, we can allow the process to take negative values, as these could only increase the drift's upper bound. However, we need to have some restriction on the step size in order to make sure not to move away from the target. Again, we provide an example (Example 9) showing this necessity at the end of this section.

Theorem 8 *(Lower additive drift, expected bounded step size). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, and let $T = \inf\{t \mid X_t \le 0\}$. Furthermore, suppose that*

(a) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \le \delta$, and that
(b) there is some value $c \ge 0$ such that, for all $t < T$, it holds that $\mathrm{E}\big[|X_{t+1} - X_t| \,\big|\, X_0, \ldots, X_t\big] \le c$.

Then

$$\mathrm{E}[T \mid X_0] \ge \frac{X_0 - \mathrm{E}[X_T \mid X_0]}{\delta} \ge \frac{X_0}{\delta} .$$

Proof. We make a case distinction with respect to $\mathrm{E}[T \mid X_0]$ being finite. If $\mathrm{E}[T \mid X_0]$ is infinite, then the theorem trivially holds. Thus, we now assume that $\mathrm{E}[T \mid X_0] < \infty$.

Similar to the proof of Theorem 4, we define, for all $t < T$, $Y_t = X_t + \delta t$, which is a submartingale, since

$$\begin{aligned} Y_t - \mathrm{E}[Y_{t+1} \mid Y_0, \ldots, Y_t] &= X_t - \delta t - \mathrm{E}[X_{t+1} - \delta(t+1) \mid X_0, \ldots, X_t] \\ &= X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] - \delta \le 0 , \end{aligned}$$

as we assume that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \le \delta$ for all $t < T$ and because, again, the transformation of X_t to Y_t is injective.

Since we now assume that $\mathrm{E}[T \mid X_0] < \infty$ and, further, that $\mathrm{E}\big[|X_{t+1} - X_t| \,\big|\, X_0, \ldots, X_t\big] \le c$ for all $t < T$, we can directly apply case (2) of Theorem 1 and get that $\mathrm{E}[Y_T \mid X_0] \ge \mathrm{E}[Y_0 \mid X_0]$. This yields, noting that $X_T \le 0$,

$$\begin{aligned} X_0 &= \mathrm{E}[X_0 \mid X_0] = \mathrm{E}[Y_0 \mid X_0] \le \mathrm{E}[Y_T \mid X_0] = \mathrm{E}[X_T + \delta T \mid X_0] \\ &= \mathrm{E}[X_T \mid X_0] + \delta \mathrm{E}[T \mid X_0] . \end{aligned}$$

Thus, for the first bound, we get the desired bound by solving for $\mathrm{E}[T \mid X_0]$. For the second bound, we use the first bound and that $\mathrm{E}[X_T \mid X_0] \le 0$. This concludes the proof. □

Note that the step size has to be bounded in some way for a lower bound, as the following example shows.

Example 9. Let $\delta \in (0, 1)$, and let $(X_t)_{t\in\mathbb{N}}$ be a random process with $X_0 = 2$ and, for all $t \in \mathbb{N}$, $X_{t+1} = 0$ with probability $1/2$ and $X_{t+1} = 2X_t - 2\delta$ otherwise. Further, let T denote the first point in time t such that $X_t = 0$. Then T follows a geometric distribution with success probability $1/2$, which yields $\mathrm{E}[T] = 2$. However, we have that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] = \delta$. If Theorem 8 could be applied to this process (by neglecting the condition of the bounded step size), the theorem would yield that $\mathrm{E}[T] \ge 2/\delta$, which is not true.

4. Variable drift

In contrast to additive drift, *variable* drift means that the drift can depend on the current state of the process (while still being bounded independently of the time). Interestingly, these more flexible drift theorems can be derived by using additive drift. Intuitively, the reasoning behind this approach is to scale the state space such that the information relevant to the current state of the process cancels out.

It is important to note that variable drift theorems are commonly phrased such that the first-hitting time T denotes the first point in time such that the random process drops *strictly below* a certain value (our target) – it is not enough to hit that value. However, this restriction is not always necessary. Thus, we also consider the setting from Section 3, where T denotes the first point in time such that we hit our target *or* get below it. In this section, our target is no longer 0 but a value $x_{\min}$.

In all of our theorems in this section, we make use of a set D. This set contains (at least) all possible values that our process can take while not having reached the target yet. It is a formal necessity in order to calculate the bound of the first-hitting time (via an integral). However, when applying the theorem, it is usually sufficient to choose $D = \mathbb{R}$ or $D = \mathbb{R}_{\ge 0}$.

4.1. Upper bounds

The first variable drift theorem was proven by Johannsen [5] and, independently in a different version, by Mitavskiy et al. [20]. It was later refined by Rowe and Sudholt [21]. In all of these versions, bounded state spaces were used. Due to Theorem 7, we can drop this restriction.

4.1.1. Going strictly below the target

The following version of the theorem assumes that the process has to drop strictly below the target, denoted by $x_{\min}$. We provide the other version afterward.

Theorem 10 *(Upper variable drift, unbounded, strictly below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that*

(a) $X_0 \geq x_{\min}$ and, for all $t \leq T$, it holds that $X_t \geq 0$ and that
(b) there is a monotonically increasing function $h\colon D \to \mathbb{R}^+$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq h(X_t)$.

Then

$$\mathrm{E}[T \mid X_0] \leq \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X_0} \frac{1}{h(z)} \mathrm{d}z \,.$$

Proof. The proof follows the one given by Rowe and Sudholt [21] very closely. We define a function $g\colon D \cup [0, x_{\min}] \to \mathbb{R}_{\geq 0}$ as follows:

$$g(x) = \begin{cases} 0 & \text{if } x < x_{\min}, \\ \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{x} \frac{1}{h(z)} \mathrm{d}z & \text{else.} \end{cases}$$

Note that g is well-defined, since $1/h$ is monotonically decreasing and every monotone function is integrable over all compact intervals of its domain. Further, $g(X_t) = 0$ holds if and only if $X_t < x_{\min}$. Thus, both processes have the same first-hitting time.

Assume that $x \geq y \geq x_{\min}$. We get

$$g(x) - g(y) = \int_{y}^{x} \frac{1}{h(z)} \mathrm{d}z \geq \frac{x-y}{h(x)} \,,$$

since h is monotonically increasing. Assuming $y \geq x \geq x_{\min}$, we get, similar to before,

$$g(x) - g(y) = -\int_{x}^{y} \frac{1}{h(z)} \mathrm{d}z \geq -\frac{y-x}{h(x)} = \frac{x-y}{h(x)} \,.$$

Thus, we can write, for $x \geq x_{\min}$ and $y \geq x_{\min}$,

$$g(x) - g(y) \geq \frac{x-y}{h(x)} \,.$$

Further, for $x \geq x_{\min} > y \geq 0$, we get

$$\begin{aligned} g(x) - g(y) &= \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{x} \frac{1}{h(z)} \mathrm{d}z \geq \frac{x_{\min}}{h(x)} + \frac{x - x_{\min}}{h(x)} \\ &= \frac{x}{h(x)} \geq \frac{x-y}{h(x)} \,. \end{aligned}$$

Overall, for $x \geq x_{\min}$ (including $X_0 \geq x_{\min}$) and $y \in \mathbb{R}_{\geq 0}$, we can estimate

$$g(x) - g(y) \geq \frac{x-y}{h(x)} \,.$$

We use this to determine the drift of the process $g(X_t)$ as follows:

$$\begin{aligned} g(X_t) - \mathrm{E}[g(X_{t+1}) \mid X_0, \ldots, X_t] &= \mathrm{E}[g(X_t) - g(X_{t+1}) \mid X_0, \ldots, X_t] \\ &\geq \frac{\mathrm{E}[X_t - X_{t+1} \mid X_0, \ldots, X_t]}{h(X_t)} \geq 1 \,, \end{aligned}$$

where we used the condition on the drift of X_t.

An application of Theorem 7 completes the proof. □

4.1.2. Going below the target

As mentioned before, it is not always necessary to drop strictly below the target. For the additive drift, for example, we are interested in the first time reaching the target. Interestingly, the proof for the following theorem is straightforward, as it is almost the same as the proof of Theorem 10. Intuitively, the waiting time for getting below the target, once it is reached, is eliminated from the expected first-hitting time. However, it is important to note that it is now not allowed to get below the target. Thus, for this bound, we actually bound the expected time until we hit the target for the first time.

Theorem 11 *(Upper variable drift, unbounded, below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} \geq 0$, and let $T = \inf\{t \mid X_t \leq x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that,*

(a) for all $t \leq T$, it holds that $X_t \geq x_{\min}$ and that
(b) there is a monotonically increasing function $h\colon D \to \mathbb{R}^+$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq h(X_t)$.

Then

$$\mathrm{E}[T \mid X_0] \leq \int_{x_{\min}}^{X_0} \frac{1}{h(z)} \mathrm{d}z\,.$$

Proof. This proof is almost identical to the proof of Theorem 10. The difference is that we define our potential function $g\colon D \to \mathbb{R}_{\geq 0}$ as follows:

$$g(x) = \begin{cases} 0 & \text{if } x \leq x_{\min}, \\ \int_{x_{\min}}^{x} \frac{1}{h(z)} \mathrm{d}z & \text{else.} \end{cases}$$

As for $g(x) - g(y)$, the case $x \geq x_{\min} > y$ does not exist anymore, since we cannot get below $x_{\min}$. Thus, the potential difference is the same in all cases, and nothing changes in the rest of the proof. □

4.2. Lower bounds

Similar to how the step size of the lower bound of the additive drift needs to be bounded, the step size for variable drift is also bounded in order to derive lower bounds. Doerr et al. [22] prove a lower-bound for variable drift (Theorem 7) that is very similar to our following theorems but a bit more strict, as it assumes that the process is monotonically decreasing over time. Our theorems only assume that the step size of the process is bounded. Further, Gießen and Witt [23] prove a variable drift theorem yielding a lower bound that is applicable to any process that does not change too much within a single step with a certain high probability (bounded by a function similar to the drift function h). We give results for more restricted processes with deterministically bounded step sizes and get easier theorems in return. As an additional constraint, we bound how fast the drift function h can grow. This can be done in various ways. We consider two cases: one case where the function bounding the drift can increase up to a multiplicative factor over a bounded range, and one case where it can only change by an additive term. As in the case of the lower bound of the additive drift theorem (Theorem 8), our process can take negative values.

Ideally, we would get the same bounds as in Section 4.1. However, we are off by a factor of $1/s$, where $s \geq 1$ is a slack term heavily depending on the parameters of the process. Thus, the quality of our lower bounds can deviate drastically between different processes. To be more precise, the slack term s depends on the maximum step size c of the random process as well as the speed of growth a of the drift function h (either relative or absolute) and on the minimal value of h. Further, the parameter c is hard to adjust when considering a given process. In addition to that, as we discuss in the following section, c may even depend on $x_{\min}$ which, in return, can lead to a bad slack term s. This means that the only sensible possibility of improving the slack term s and thus the lower bound of the expected first-hitting time is to come up with a good drift function h, which influences the minimal value of h as well as the speed of growth a.

4.2.1. Going strictly below the target

In this scenario, we require our process to not take values in the interval $(0, x_{\min})$ – the *gap*. Further, as mentioned above, we always require our random process to have a (uniformly) bounded step size c. It is very important to stress that the gap impacts c, as the process has to be able to get past the gap; in return, c influences the quality of the lower bound drastically, as it impacts the slack term s. Thus, if it is possible to consider the first point in time that the target can be *hit*, we advise using our theorems given in Section 4.2.2, as they do not have this problem.

Theorem 12 *(Lower variable drift, bounded step size, relative difference, strictly below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that*

(a) $X_0 \geq x_{\min}$ *and, for all* $t \leq T$, $X_t \notin (0, x_{\min})$, *that*
(b) *there is some value* $c \geq 0$ *such that, for all* $t < T$, *it holds that* $|X_{t+1} - X_t| \leq c$, *and that*
(c) *there is a monotonically increasing function* $h\colon D \cup [x_{\min} - c, x_{\min}] \to \mathbb{R}^+$ *such that,*
- *there is some value* $a \geq 1$ *such that* $h(x+c) \leq a \cdot h(x-c)$ *and that,*
- *for all* $t < T$, *we have* $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \leq h(X_t)$.

Let $s = 1 + \frac{(a-1)c}{a \cdot h(x_{\min}-c)}$. *Then*

$$\mathrm{E}[T \mid X_0] \geq \frac{1}{s} \cdot \left(\frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X_0} \frac{1}{h(z)} \mathrm{d}z \right) .$$

Proof. Similar to the proof of Theorem 10, we define a function $g\colon D \cup [x_{\min} - c, x_{\min}] \to \mathbb{R}_{\geq 0}$ as follows:

$$g(x) = \begin{cases} 0 & \text{if } x < x_{\min}, \\ \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{x} \frac{1}{h(z)} \mathrm{d}z & \text{else.} \end{cases}$$

Again, $g(X_t) = 0$ if and only if $X_t \leq 0$.

Assume that $x \geq y \geq x_{\min}$ with $x - y \leq c$. Thus, $y \geq x - c$. We get

$$g(x) - g(y) = \int_{y}^{x} \frac{1}{h(z)} \mathrm{d}z \leq \frac{x-y}{h(y)} \leq \frac{x-y}{h(x-c)} ,$$

since h is monotonically increasing. Assuming $y \geq x \geq x_{\min}$ with $y - x \leq c$ and noting that $y \leq x + c$, we obtain,

$$g(x) - g(y) = -\int_{x}^{y} \frac{1}{h(z)} \mathrm{d}z \leq -\frac{y-x}{h(y)} \leq \frac{x-y}{h(x+c)} ,$$

similar to before. Further, for $x \geq x_{\min} > 0 \geq y$ with $x - x_{\min} < c$, implying that $x_{\min} > x - c$, we get

$$\begin{aligned} g(x) - g(y) &= \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{x} \frac{1}{h(z)} \mathrm{d}z \leq \frac{x_{\min}}{h(x_{\min})} + \frac{x - x_{\min}}{h(x_{\min})} \\ &= \frac{x}{h(x_{\min})} \leq \frac{x-y}{h(x_{\min})} \leq \frac{x-y}{h(x-c)} . \end{aligned}$$

For a predicate P, let $[P]$ denote the Iverson bracket, i.e., the characteristic function with respect to P. We determine the drift of $g(X_t)$ as follows, making use of the previous inequalities:

$$\begin{aligned} & g(X_t) - \mathrm{E}[g(X_{t+1}) \mid X_0, \ldots, X_t] \\ &\quad = \mathrm{E}\big[\big(g(X_t) - g(X_{t+1})\big)[X_t > X_{t+1}] \,\big|\, X_0, \ldots, X_t\big] \\ &\qquad + \mathrm{E}\big[\big(g(X_t) - g(X_{t+1})\big)[X_t < X_{t+1}] \,\big|\, X_0, \ldots, X_t\big] \\ &\quad \leq \frac{1}{h(X_t - c)} \mathrm{E}\big[(X_t - X_{t+1})[X_t > X_{t+1}] \,\big|\, X_0, \ldots, X_t\big] \\ &\qquad + \frac{1}{h(X_t + c)} \mathrm{E}\big[(X_t - X_{t+1})[X_t < X_{t+1}] \,\big|\, X_0, \ldots, X_t\big], \end{aligned}$$

where the term $\mathrm{E}\big[(X_t - X_{t+1})[X_t < X_{t+1}] \,\big|\, X_0, \ldots, X_t\big]$ is negative. Using the assumption $h(x+c) \leq a \cdot h(x-c)$, we can thus upper-bound the previous term and obtain

$$\begin{aligned} & g(X_t) - \mathrm{E}[g(X_{t+1}) \mid X_0, \ldots, X_t] \\ & \leq \frac{1}{h(X_t - c)} \mathrm{E}\big[(X_t - X_{t+1})[X_t > X_{t+1}] \,\big|\, X_0, \ldots, X_t\big] \\ & + \frac{1}{a \cdot h(X_t - c)} \mathrm{E}\big[(X_t - X_{t+1})[X_t < X_{t+1}] \,\big|\, X_0, \ldots, X_t\big] \\ & = \frac{a-1}{a \cdot h(X_t - c)} \mathrm{E}\big[(X_t - X_{t+1})[X_t > X_{t+1}] \,\big|\, X_0, \ldots, X_t\big] \end{aligned}$$

$$+\frac{1}{a\cdot h(X_t-c)}\Big(\mathrm{E}\big[(X_t-X_{t+1})[X_t>X_{t+1}]\,\big|\,X_0,\ldots,X_t\big] + \mathrm{E}\big[(X_t-X_{t+1})[X_t<X_{t+1}]\,\big|\,X_0,\ldots,X_t\big]\Big)$$
$$\le \frac{(a-1)c}{a\cdot h(X_t-c)}+\frac{1}{a\cdot h(X_t-c)}\mathrm{E}[X_t-X_{t+1}\mid X_0,\ldots,X_t],$$

where the last inequality made use of $|X_t-X_{t+1}|\le c$.

Using our assumption $X_t-\mathrm{E}[X_{t+1}\mid X_0,\ldots,X_t]\le h(X_t)$, we finally get

$$\begin{aligned} g(X_t)-\mathrm{E}[g(X_{t+1})\mid X_0,\ldots,X_t] &\le \frac{(a-1)c}{a\cdot h(X_t-c)}+\frac{h(X_t)}{a\cdot h(X_t-c)}\\ &\le \frac{(a-1)c}{a\cdot h(x_{\min}-c)}+\frac{h(X_t)}{h(X_t+c)}\\ &\le \frac{(a-1)c}{a\cdot h(x_{\min}-c)}+1=s\,. \end{aligned}$$

Applying Theorem 8 completes the proof. □

Note how a constant drift function h (that is, $a=1$) yields, in combination with Theorem 10, a tight run time bound. Further, if $c=\mathrm{O}\big(h(x_{\min}-c)\big)$, then s becomes constant and the bound is tight up to a constant factor. In general, the higher $h(x_{\min}-c)$ or the smaller c, the smaller s and thus the better the bound.

We can state Theorem 12 in a slightly different fashion by restricting by how much $h(x+c)$ and $h(x-c)$ may differ, resulting in a more complicated but also more slowly growing slack term s.

Corollary 13 *(Lower variable drift, bounded step size, absolute difference, strictly below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min}>0$, and let $T=\inf\{t\mid X_t<x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x\ge x_{\min}$ that, for all $t\le T$, any X_t can take. Furthermore, suppose that*

(a) $X_0\ge x_{\min}$ and, for all $t\le T$, $X_t\notin(0,x_{\min})$, that
(b) there is some value $c\ge 0$ such that, for all $t<T$, it holds that $|X_{t+1}-X_t|\le c$, and that
(c) there is a monotonically increasing function $h\colon D\cup[x_{\min}-c,x_{\min}]\to\mathbb{R}^+$ such that
- *there is some value $a\ge 0$ such that $h(x+c)\le a+h(x-c)$ and that,*
- *for all $t<T$, we have $X_t-\mathrm{E}[X_{t+1}\mid X_0,\ldots,X_t]\le h(X_t)$.*

Let $s=1+\frac{ac}{\left(a+h(x_{\min}-c)\right)h(x_{\min}-c)}$. Then

$$\mathrm{E}[T\mid X_0]\ge\frac{1}{s}\cdot\left(\frac{x_{\min}}{h(x_{\min})}+\int_{x_{\min}}^{X_0}\frac{1}{h(z)}\mathrm{d}z\right).$$

Proof. We want to apply Theorem 12. We define

$$a'=\frac{a}{h(x_{\min}-c)}+1\,.$$

Note that $a\ge 1$. Further, for all $x\in D$,

$$\begin{aligned} h(x+c)\le a+h(x-c) &= (a'-1)h(x_{\min}-c)+h(x_{\min}-c)\\ &\le a'\cdot h(x_{\min}-c)\le a'\cdot h(x-c)\,, \end{aligned}$$

as h is monotone. Thus, we can apply Theorem 12 and get

$$s=1+\frac{(a'-1)c}{a'\cdot h(x_{\min}-c)}=1+\frac{\frac{a}{h(x_{\min}-c)}}{\left(\frac{a}{h(x_{\min}-c)}+1\right)h(x_{\min}-c)}\,,$$

which concludes the proof. □

4.2.2. Going below the target

The following theorems are versions of the last two theorems. However, now the first-hitting time denotes the first time such that the random process at least reaches the target. This removes the necessity of the gap from the previous theorems, yielding better lower bounds in return. Further, since the step width c is no longer tied to a gap, the slack term s may be better.

Theorem 14 *(Lower variable drift, bounded step size, relative difference, below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} \geq 0$, and let $T = \inf\{t \mid X_t \leq x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that*

(a) $X_0 \geq x_{\min}$, that
(b) there is some value $c \geq 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \leq c$, and that
(c) there is a monotonically increasing function $h\colon D \cup [x_{\min} - c, x_{\min}] \to \mathbb{R}^+$ such that
- *there is some value $a \geq 0$ such that $h(x + c) \leq a \cdot h(x - c)$ and that,*
- *for all $t < T$, we have $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \leq h(X_t)$.*

Let $s = 1 + \frac{(a-1)c}{a\cdot h(x_{\min}-c)}$. Then

$$\mathrm{E}[T \mid X_0] \geq \frac{1}{s} \cdot \int_{x_{\min}}^{X_0} \frac{1}{h(z)} \mathrm{d}z \,.$$

Proof. This proof is almost identical to the proof of Theorem 12. The difference is that we define our potential function $g\colon D \cup [x_{\min} - c, x_{\min}] \to \mathbb{R}_{\geq 0}$ as follows:

$$g(x) = \begin{cases} 0 & \text{if } x \leq x_{\min}, \\ \int_{x_{\min}}^{x} \frac{1}{h(z)} \mathrm{d}z & \text{else.} \end{cases}$$

As for $g(x) - g(y)$, we need to reconsider the case $x \geq x_{\min} > 0 \geq y$, which now translates to $x \geq x_{\min} \geq y$. We get

$$g(x) - g(y) = \int_{x}^{x_{\min}} \frac{1}{h(z)} \mathrm{d}z \leq \frac{x - x_{\min}}{h(x_{\min})} \leq \frac{x - x_{\min}}{h(x-c)} \leq \frac{x - y}{h(x-c)} \,,$$

where the last inequality is due to $y \leq x_{\min}$.

In all other cases, the previously used term $x_{\min}/h(x_{\min})$ canceled out. Thus, the potential difference is the same in all cases, and nothing changes in the rest of the proof. □

As in Section 4.2.1, we get a better slack term by demanding a stricter restriction on the speed of growth of h.

Corollary 15 *(Lower variable drift, bounded step size, absolute difference, below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} \geq 0$, and let $T = \inf\{t \mid X_t \leq x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that*

(a) $X_0 \geq x_{\min}$, that
(b) there is some value $c \geq 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \leq c$, and that
(c) there is a monotonically increasing function $h\colon D \cup [x_{\min} - c, x_{\min}] \to \mathbb{R}^+$ such that
- *there is some value $a \geq 0$ such that $h(x + c) \leq a + h(x - c)$ and that,*
- *for all $t < T$, we have $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \leq h(X_t)$.*

Let $s = 1 + \frac{ac}{(a+h(x_{\min}-c))h(x_{\min}-c)}$. Then

$$\mathrm{E}[T \mid X_0] \geq \frac{1}{s} \int_{x_{\min}}^{X_0} \frac{1}{h(z)} \mathrm{d}z \,.$$

Proof. This proof changes only in the same places as the proof of Theorem 14. This does not change the line of argument of the proof of Corollary 13, which concludes this proof. □

5. Multiplicative drift

A special case of variable drift is *multiplicative* drift, where the drift can be bounded by a multiple of the most recent value in the history of the process. As before, we provide upper and lower bounds in the two versions of either dropping strictly below the target or permitting to hit it. In this setting, it can be intuitively argued why the version of dropping strictly below the target is useful: consider a sequence of nonnegative numbers that halves its current value each time step. This process will never reach 0 within finite time. However, it drops below any value greater than 0.

5.1. Upper bounds

Both upper bounds we state are simple applications of the corresponding variable drift theorems from Section 4.1.

5.1.1. Going strictly below the target

Corollary 16 has first been stated by Doerr et al. [6] using finite state spaces. Afterward, it has been proven multiple times for processes not requiring an upper bound (although this is not always stated) [10–12].

Corollary 16 *(Upper multiplicative drift, unbounded, strictly below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Furthermore, suppose that*

(a) $X_0 \geq x_{\min}$ and, for all $t \leq T$, it holds that $X_t \geq 0$, and that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}\big[X_{t+1} \,\big|\, X_0, \ldots, X_t\big] \geq \delta X_t$.

Then

$$\mathrm{E}[T \mid X_0] \leq \frac{1 + \ln\left(\frac{X_0}{x_{\min}}\right)}{\delta} .$$

Proof. We define a function $h\colon [x_{\min}, \infty) \to \mathbb{R}^+$ with $h(x) = \delta x$. Note that h is monotonically increasing and that, by construction, for all $t < T$, $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \geq h(X_t)$. Thus, by applying Theorem 10, we get

$$\mathrm{E}[T \mid X_0] \leq \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X_0} \frac{1}{h(z)}\mathrm{d}z = \frac{x_{\min}}{\delta x_{\min}} + \frac{\ln\left(\frac{X_0}{x_{\min}}\right)}{\delta} ,$$

which completes the proof. □

5.1.2. Going below the target

By applying Theorem 11 instead of Theorem 10, we get the following theorem. As in the case of Theorem 11, the process now has to be lower-bounded by $x_{\min}$.

Corollary 17 *(Upper multiplicative drift, unbounded, below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t \leq x_{\min}\}$. Furthermore, suppose that,*

(a) for all $t \leq T$, it holds that $X_t \geq x_{\min}$, and that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}\big[X_{t+1} \,\big|\, X_0, \ldots, X_t\big] \geq \delta X_t$.

Then

$$\mathrm{E}[T \mid X_0] \leq \frac{\ln\left(\frac{X_0}{x_{\min}}\right)}{\delta} .$$

Proof. We define the same potential as in the proof of Corollary 16 but apply Theorem 11 instead. □

Before we consider lower bounds, we want to provide an example that shows that the upper bounds are as tight as possible, up to constant factors, for the range of processes we consider. The example describes a process that decreases deterministically, that is, it has a variance of 0. Interestingly, for the lower bounds, we provide an example process (Example 24) with maximal variance (that still has a positive drift) which shows that our lower bounds are tight.

Example 18. Let $\delta \in (0, 1)$ be a value bounded away from 1. Consider the process $(X_t)_{t\in\mathbb{N}}$, with $X_0 > 1$, that decreases each step deterministically such that $X_{t+1} = (1 - \delta)X_t$ holds. Let T denote the first point in time such that the process drops below 1. Thus, we get $T = \Theta(-\log_{(1-\delta)} X_0) = \Theta\big(-\ln(X_0)/\ln(1-\delta)\big) = \Theta\big(\ln(X_0)/\delta\big)$, where the last equation makes use of the Taylor expansion of $\ln(1-\delta) = \Theta(-\delta)$, as $1-\delta$ does not converge to 1, by assumption.

5.2. Lower bounds

In this section, we first give the multiplicative versions that follow from the respective variable versions from Section 4.2. However, as we already discussed in Section 4.2.1, the lower bounds can be bad, depending on the choice of $x_{\min}$, as this influences the slack term s. In order to mitigate this effect, Section 5.2.3 provides two other theorems that give better lower bounds with respect to where the random process starts and when we stop.

5.2.1. Going strictly below the target

Similar to the upper bound before, we now use Corollary 13 in order to state a lower bound formulation of the multiplicative drift theorem.

Corollary 19 *(Lower multiplicative drift, bounded step size, strictly below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Furthermore, suppose that*

(a) $X_0 \geq x_{\min}$ and, for all $t \leq T$, $X_t \notin (0, x_{\min})$, that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \leq \delta X_t$, and that
(c) there is some value $c \geq 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \leq c$.

Let $s = 1 + \frac{2c^2}{(2c+x_{\min})\delta x_{\min}}$. Then

$$\mathrm{E}[T \mid X_0] \geq \frac{1}{s} \cdot \frac{1 + \ln\left(\frac{X_0}{x_{\min}}\right)}{\delta}.$$

Proof. We define a function $h\colon [x_{\min} - c, \infty) \to \mathbb{R}^+$ as follows:

$$h(x) = \begin{cases} \delta x & \text{if } x \geq x_{\min}, \\ \delta x_{\min} & \text{else.} \end{cases}$$

Note that h is monotonically increasing. Further, it holds that $h(x + c) \leq 2\delta c + h(x - c)$ and that, for all $t < T$, $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \leq h(X_t)$. Thus, we can apply Corollary 13 with $a = 2\delta c$. We get

$$s = 1 + \frac{ac}{\big(a + h(x_{\min} - c)\big)h(x_{\min} - c)} = 1 + \frac{2\delta c^2}{(2\delta c + \delta x_{\min})\delta x_{\min}} = 1 + \frac{2c^2}{(2c + x_{\min})\delta x_{\min}}$$

and, thus,

$$\mathrm{E}[T \mid X_0] \geq \frac{1}{s} \cdot \left(\frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X_0} \frac{1}{h(z)} \mathrm{d}z\right) = \frac{1}{s} \cdot \left(\frac{x_{\min}}{\delta x_{\min}} + \frac{\ln\left(\frac{X_0}{x_{\min}}\right)}{\delta}\right) = \frac{1}{s} \cdot \frac{1 + \ln\left(\frac{X_0}{x_{\min}}\right)}{\delta},$$

which is what we claimed. □

Note that a similar corollary can be obtained by using Theorem 12. However, if done so, the result is not as good.

We want to mention that Corollary 19 is very similar to Thm. 3.3 from Doerr et al. [24], which stands by itself, whereas our corollary follows from our more general Corollary 13. In comparison, the slack term of Doerr et al. [24] is $2 + c^2/\big(\delta(x_{\min}^2 - c^2)\big)$.

Note that the lower bound of Corollary 19 differs from the upper bound of Corollary 16 only by the factor $1/s$. However, if $c = \Theta(x_{\min})$ and $\delta = O(1)$, then $s\delta = \Theta(1)$, which results in the gap between the upper and the lower bound being in the order of δ.

As we already discussed in Section 4.2, it is basically impossible to adjust c, as this is an inherent property of the random process X. However, the choice of $x_{\min}$ usually is flexible, allowing to adjust s. Interestingly, the smaller $x_{\min}$, the worse the lower bound becomes. Thus, choosing $x_{\min}$ reasonably large can yield satisfactory results. Unfortunately, increasing $x_{\min}$ may also entail increasing c when a gap is present. Thus, the following corollary is better suited for this kind of analysis.

5.2.2. Going below the target

Corollary 20 *(Lower multiplicative drift, bounded step size, below target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t \le x_{\min}\}$. Furthermore, suppose that*

(a) $X_0 \ge x_{\min}$, that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \le \delta X_t$, and that
(c) there is some value $c \ge 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \le c$.

Let $s = 1 + \frac{2c^2}{(2c+x_{\min})\delta x_{\min}}$. Then

$$\mathrm{E}[T \mid X_0] \ge \frac{1}{s} \cdot \frac{\ln\left(\frac{X_0}{x_{\min}}\right)}{\delta} .$$

Proof. We follow the proof of Theorem 19 but apply Corollary 15 instead of Corollary 13. □

5.2.3. Different regimes

As discussed before, we now make a case distinction with respect to $x_{\min}$ (and, thus, implicitly X_0). If it is rather large, Corollary 20 already yields a reasonable bound, which we phrase as Corollary 21. However, this bound may become very bad for smaller values. This is why we further provide Theorem 23, which copes with that case.

Corollary 21 *(Lower multiplicative drift, bounded step size, below target, large regime). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $d > 0$, and let $T = \inf\{t \mid X_t \le d/\sqrt{\delta}\}$. Furthermore, suppose that*

(a) $X_0 \ge d/\sqrt{\delta}$, that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \le \delta X_t$, and that
(c) there is some value $c \ge 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \le c$.

Let $s = 1 + \frac{2c^2}{2cd\sqrt{\delta}+d^2}$. Then

$$\mathrm{E}[T \mid X_0] \ge \frac{1}{s} \cdot \frac{\ln\left(\frac{X_0\sqrt{\delta}}{d}\right)}{\delta} .$$

Proof. We apply Corollary 20 with $x_{\min} = d/\sqrt{\delta}$ and see that

$$s = 1 + \frac{2c^2}{(2c + x_{\min})\delta x_{\min}} = 1 + \frac{2c^2}{\left(2c + \frac{d}{\sqrt{\delta}}\right)\delta \frac{d}{\sqrt{\delta}}} = 1 + \frac{2c^2}{2cd\sqrt{\delta} + d^2} ,$$

which finishes the proof. □

Note that, if c and d are in $\Theta(1)$ and $\delta = O(1)$, then Corollary 21 yields a bound of $\Omega\big(\ln(X_0\sqrt{\delta})/\delta\big)$, which is tight when compared to Corollary 17 with $x_{\min} = d/\sqrt{\delta}$.

Example 22. Applied to the well-known Coupon Collector problem, which is commonly used as an example for an application of the multiplicative drift theorem, we see that only missing $\sqrt{n}$ coupons when starting with n missing coupons takes, in expectation, at least $(1/6)n\ln n$ tries. This is complemented by the corresponding upper bound $(1/2)n\log n$ from Corollary 17. Thus, the bound is tight up to constant factors. More generally speaking, if X_0 is large, a sufficiently large $x_{\min}$ suffices to get a lower bound that matches the upper up to constant factors.

We now consider the case that we start at lower values. Still in the regime of multiplicative drift of δ, we are interested in the behavior of such a process when $X_0 < d/\sqrt{\delta}$, that is, the setting we did not analyze earlier. As it turns out, the process may behave a bit different then.

Intuitively, for such small values, the first-hitting time of the process may be mainly determined by its random-walk behavior, as the multiplicative impact of the drift is not too large anymore. We make this explicit in the proof by only analyzing the first-hitting time of a submartingale of the original process, which can be seen as the process's random-walk behavior. In order to get a bound on this behavior, we need an extra restriction (condition (e)).

Theorem 23 *(Lower multiplicative drift, bounded step size, below target, small regime). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Furthermore, suppose that,*

(a) for all $t \le T$, it holds that $X_t \notin (0, x_{\min})$, that
(b) there is some value $\delta > 0$ such that, for all $t < T$, it holds that $X_t - \mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \le \delta X_t$, that
(c) there is some value $c \ge 0$ such that, for all $t < T$, it holds that $|X_{t+1} - X_t| \le c$, that
(d) there is some value $d > 0$ such that it holds that $X_0 \le d/\sqrt{\delta}$, and that
(e) there is some value $a \in \mathbb{R}$ such that, for all $t < T$, it holds that $\mathrm{E}\left[X_{t+1}^2 \,\middle|\, X_0, \ldots, X_t\right] - {X_t}^2 \le a$.

Then

$$\mathrm{E}[T \mid X_0] \ge \frac{d + c\sqrt{\delta}}{4d^2 + 4cd\sqrt{\delta} + c^2\delta + a} \cdot \frac{X_0}{\sqrt{\delta}} .$$

Proof. In the regime of this theorem, the multiplicative drift is not very large, since the values X_t themselves are rather small. Thus, the variance of the process introduced by its random walk nature can be factored in in order to achieve a better lower run time bound.

We only consider the time T' it takes X_t until it drops below $x_{\min}$ or reaches values of at least $2d/\sqrt{\delta} + c$. Note that T dominates T', as the latter has one extra target.

We define, for all $t \le T'$:

$$Y_t = X_t \left(2\frac{d}{\sqrt{\delta}} + c - X_t\right) ,$$

which is positive as long as $X_t \in [x_{\min}, 2d/\sqrt{\delta} + c)$.

When calculating the drift of Y_t with respect to X_t, we get

$$\begin{aligned} &Y_t - \mathrm{E}[Y_{t+1} \mid X_0, \ldots, X_t] \\ &\quad = X_t \left(2\frac{d}{\sqrt{\delta}} + c - X_t\right) - \mathrm{E}\left[X_{t+1}\left(2\frac{d}{\sqrt{\delta}} + c - X_{t+1}\right) \,\middle|\, X_0, \ldots, X_t\right] \\ &\quad = X_t \left(2\frac{d}{\sqrt{\delta}} + c\right) - {X_t}^2 - \left(2\frac{d}{\sqrt{\delta}} + c\right)\mathrm{E}[X_{t+1} \mid X_0, \ldots, X_t] \\ &\qquad + \mathrm{E}\left[X_{t+1}^2 \,\middle|\, X_0, \ldots, X_t\right] . \end{aligned}$$

By using assumptions (b) and (e), we get

$$\begin{aligned} Y_t - \mathrm{E}[Y_{t+1} \mid X_0, \ldots, X_t] &\le \delta X_t \left(2\frac{d}{\sqrt{\delta}} + c\right) + \mathrm{E}\left[X_{t+1}^2 \,\middle|\, X_0, \ldots, X_t\right] - {X_t}^2 \\ &\le \delta X_t \left(2\frac{d}{\sqrt{\delta}} + c\right) + a \le \delta\left(2\frac{d}{\sqrt{\delta}} + c\right)^2 + a \\ &= 4d^2 + 4cd\sqrt{\delta} + c^2\delta + a . \end{aligned}$$

Thus, by applying Theorem 8, we get,

$$\begin{aligned} \mathrm{E}\left[T' \,\middle|\, X_0\right] &\ge \frac{Y_0}{4d^2 + 4cd\sqrt{\delta} + c^2\delta + a} = \frac{X_0\left(2\frac{d}{\sqrt{\delta}} + c - X_0\right)}{4d^2 + 4cd\sqrt{\delta} + c^2\delta + a} \\ &\ge \frac{X_0\left(\frac{d}{\sqrt{\delta}} + c\right)}{4d^2 + 4cd\sqrt{\delta} + c^2\delta + a} , \end{aligned}$$

as $X_0 \le d/\sqrt{\delta}$. Since $\mathrm{E}[T \mid X_0] \ge \mathrm{E}\left[T' \,\middle|\, X_0\right]$, this concludes the proof. □

Note that, if a, c, and d are in $\Theta(1)$, $a \ge 0$, and if $\delta = \mathrm{O}(1)$, then Theorem 23 yields a bound of $\Omega\left(X_0/\sqrt{\delta}\right)$. Further note that the bound given in Theorem 23 may be asymptotically lower than the upper bounds given in Section 5.1. We want to show with the following example that there are processes where the bound given in Theorem 23 is tight up to constant factors.

Example 24. Let $\delta \in (0,1)$ be a value such that $1/\sqrt{\delta}$ is an integer. We define the following process $(X_t)_{t\in\mathbb{N}}$ over $\mathbb{N}$, which is upper-bounded by $1/\sqrt{\delta}$. If $0 < X_t < 1/\sqrt{\delta}$, then $X_{t+1} = X_t - 1$ with probability $(1+\delta X_t)/2$ and $X_{t+1} = X_t + 1$ with probability $(1-\delta X_t)/2$. If $X_t = 1/\sqrt{\delta}$, then $X_{t+1} = X_t$ with probability $1-\delta X_t$ and $X_{t+1} = X_t - 1$ with probability δX_t. Last, if $X_t = 0$, then $X_{t+1} = 0$. Further, let T denote the first point in time such that the process hits 0.

This process is dominated by the following process $(Y_t)_{t\in\mathbb{N}}$ with $Y_0 = X_0$: if $0 < Y_t < 1/\sqrt{\delta}$, then $Y_{t+1} = Y_t \pm 1$ both with probability 1/2. If $Y_t = 1/\sqrt{\delta}$, then $Y_{t+1} = Y_t$ with probability $1-\sqrt{\delta}$, and $Y_{t+1} = Y_t - 1$ with probability $\sqrt{\delta}$. Last, if $Y_t = 0$, then $Y_{t+1} = 0$. Note that T is upper-bounded by the first time t such that $Y_t = 0$.

By standard arguments over unbiased random walks, we get an upper bound of $\mathrm{O}\big(Y_0/\sqrt{\delta}\big) = \mathrm{O}\big(X_0/\sqrt{\delta}\big)$ for the time until the process terminates.

6. Drift without drift

In order for a drift theorem to be applicable, the process needs to have a positive drift. However, sometimes one is interested in the first-hitting time of unbiased processes, that is, processes with a drift 0. The classical example for that is the Gambler's Ruin process, which describes a fair random walk.

In this section, we focus on such unbiased processes, that is, martingales. We show that in these cases the variance (which is nonnegative by definition) can be used in order to apply a drift theorem. Since the variance of a process is 0 if and only if the process is deterministic, we get a framework applicable to any unbiased random process.

We start by providing a transformation of a martingale into another random process that has positive drift. The underlying method of this transformation is known as *predictable quadratic variation*, although it is not always referred to under this name. For more information, see, for example, the books of Durrett [18, Chapter 4.5] or Williams [25, Chapter 12.11].

Theorem 25 *(Martingale drift transformation). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, let $[\alpha,\beta] \subset \mathbb{R}$ be an interval, and let $T = \inf\{t \in \mathbb{N} \mid X_t \notin (\alpha,\beta)\}$. Furthermore, suppose that,*

(a) for all $t < T$, it holds that $\mathrm{E}[X_{t+1} \mid X_0,\dots,X_t] = X_t$ and that,
(b) for all $t < T$, it holds that $\mathrm{Var}[X_{t+1} \mid X_0,\dots,X_t] > 0$.

Then the process $(Y_t)_{t\in\mathbb{N}}$ with

$$Y_t = (X_t - \alpha)(\beta - X_t)$$

is, for all $t < T$, a random process with positive drift $\mathrm{Var}[X_{t+1} \mid X_0,\dots,X_t]$ toward 0.

Proof. For all $t < T$, we determine the drift of Y_t with respect to X_t[3]:

$$\begin{aligned}
&\mathrm{E}[Y_t - Y_{t+1} \mid X_0,\dots,X_t] = Y_t - \mathrm{E}[Y_{t+1} \mid X_0,\dots,X_t]\\
&= (X_t-\alpha)(\beta-X_t) - \mathrm{E}[(X_{t+1}-\alpha)(\beta-X_{t+1}) \mid X_0,\dots,X_t]\\
&= -X_t^2 + (\alpha+\beta)X_t - \alpha\beta - \mathrm{E}\Big[-X_{t+1}^2 + (\alpha+\beta)X_{t+1} - \alpha\beta \,\Big|\, X_0,\dots,X_t\Big]\\
&= -X_t^2 + \mathrm{E}\Big[X_{t+1}^2 \,\Big|\, X_0,\dots,X_t\Big] + (\alpha+\beta)X_t\\
&\quad - (\alpha+\beta)\overbrace{\mathrm{E}[X_{t+1} \mid X_0,\dots,X_t]}^{=X_t} - \alpha\beta + \alpha\beta\\
&= \mathrm{E}\Big[X_{t+1}^2 \,\Big|\, X_0,\dots,X_t\Big] - \overbrace{X_t}^{=\mathrm{E}[X_{t+1} \mid X_0,\dots,X_t]}{}^2\\
&= \mathrm{E}\Big[X_{t+1}^2 \,\Big|\, X_0,\dots,X_t\Big] - \mathrm{E}[X_{t+1} \mid X_0,\dots,X_t]^2\\
&= \mathrm{Var}[X_{t+1} \mid X_0,\dots,X_t]\,,
\end{aligned}$$

which is positive by assumption. □

Note that the transformed process Y_t described in Theorem 25 is positive as long as $X_t \in (\alpha,\beta)$, and nonpositive otherwise. This means that T also denotes the first-hitting time of $Y_t \le 0$. Hence, Y_t can be used in order to apply any drift theorem where the target should be hit.

[3] We use here that our drift theorems can be used for any filtration that the process is adapted to. Note that Y_t is adapted to the natural filtration of X_t because knowing X_t fully determines Y_t.

Corollary 26 *(Martingale upper additive drift). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $[\alpha,\beta]\subset\mathbb{R}$, and let $T=\inf\{t\in\mathbb{N}\mid X_t\in\{\alpha,\beta\}\}$. Furthermore, suppose that,*

(a) for all $t<T$, it holds that $\mathrm{E}[X_{t+1}\mid X_0,\dots,X_t]=X_t$ and that
(b) there is some value $\delta>0$ such that, for all $t<T$, it holds that $\mathrm{Var}[X_{t+1}\mid X_0,\dots,X_t]\geq\delta$.

Then

$$\mathrm{E}[T\mid X_0]\leq\frac{(X_0-\alpha)(\beta-X_0)}{\delta}\,.$$

Proof. We use Theorem 25 to transform X_t into the process Y_t, which has a drift of at least δ, by assumption. Note that, for all $t\leq T$, it holds that $Y_t\geq 0$. Applying Theorem 7 completes the proof. □

For the lower bound, the martingale itself does not have to be bounded but only the state space. Due to the boundedness of the state space, we do not require a restriction on the expected step size.

Corollary 27 *(Martingale lower additive drift). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $\mathbb{R}$, let $[\alpha,\beta]\subset\mathbb{R}$ be an interval, and let $T=\inf\{t\in\mathbb{N}\mid X_t\notin(\alpha,\beta)\}$. Furthermore, suppose that,*

(a) for all $t<T$, it holds that $\mathrm{E}[X_{t+1}\mid X_0,\dots,X_t]=X_t$ and that
(b) there is some value $\delta>0$ such that, for all $t<T$, it holds that $\mathrm{Var}[X_{t+1}\mid X_0,\dots,X_t]\leq\delta$.

Then

$$\mathrm{E}[T\mid X_0]\geq\frac{(X_0-\alpha)(\beta-X_0)}{\delta}\,.$$

Proof. We use Theorem 25 and want to apply Theorem 8. For this, we can argue analogously as in the proof of Theorem 26. However, we still need to check the expected bounded step size of Y_t.

Note that, for all $t<T$, we have $\alpha\leq X_t\leq\beta$. Hence, the convex function Y_t is maximal for $X_t=(\alpha+\beta)/2$, resulting in $|Y_t|\leq\big((\alpha+\beta)/2\big)^2$. Thus, in order to bound

$$\mathrm{E}[|Y_{t+1}-Y_t|\mid X_0,\dots,X_t]\leq\mathrm{E}[|Y_{t+1}|\mid X_0,\dots,X_t]+|Y_t|$$

we are left with bounding

$$\begin{aligned}\mathrm{E}[|Y_{t+1}|\mid X_0,\dots,X_t]&=\mathrm{E}\big[|(X_{t+1}-\alpha)(\beta-X_{t+1})|\,\big|\,X_0,\dots,X_t\big]\\&\leq|\alpha+\beta|\mathrm{E}\big[|X_{t+1}|\,\big|\,X_0,\dots,X_t\big]+\mathrm{E}\Big[X_{t+1}^2\,\Big|\,X_0,\dots,X_t\Big]+|\alpha\beta|\,.\end{aligned}$$

Since we can bound

$$\mathrm{Var}[X_{t+1}\mid X_0,\dots,X_t]=\mathrm{E}\Big[X_{t+1}^2\,\Big|\,X_0,\dots,X_t\Big]-\mathrm{E}[X_{t+1}\mid X_0,\dots,X_t]^2$$

by assumption, we can bound the two expected values $\mathrm{E}\big[X_{t+1}^2\,\big|\,X_0,\dots,X_t\big]$ and $\mathrm{E}[|X_{t+1}|\mid X_0,\dots,X_t]$ and therefore $\mathrm{E}[|Y_{t+1}-Y_t|\mid X_0,\dots,X_t]$. Applying Theorem 8 finishes the proof. □

The other drift theorems follow analogously, using Theorem 25, albeit getting more complicated. As an example, we state the variable drift theorem for martingales that follows from applying Theorem 11.

Corollary 28 *(Martingale upper variable drift, hitting target). Let $(X_t)_{t\in\mathbb{N}}$ be random variables over $[\alpha,\beta]\subset\mathbb{R}$, and let $T=\inf\{t\in\mathbb{N}\colon X_t\in\{\alpha,\beta\}\}$. Furthermore, suppose that,*

(a) for all $t<T$, it holds that $\mathrm{E}[X_{t+1}\mid X_0,\dots,X_t]=X_t$ and that
(b) there is a monotonically increasing function $h\colon[0,(\alpha+\beta)^2/4]\to\mathbb{R}^+$ such that, for all $t<T$, we have $\mathrm{Var}[X_{t+1}\mid X_0,\dots,X_t]\geq h\big((X_t-\alpha)(\beta-X_t)\big)$.

Then

$$\mathrm{E}[T\mid X_0]\leq\int_0^{(X_0-\alpha)(\beta-X_0)}\frac{1}{h(z)}\mathrm{d}z\,.$$

As we mentioned at the beginning of this section, Theorem 25 allows to generalize from processes like the Gambler's Ruin. In the following, we want to use some of the above theorems in order to get the exact first-hitting time of said process.

Example 29 *(Gambler's ruin).* Let $n \in \mathbb{N}$, and let $(X_t)_{t \in \mathbb{N}}$ be a random process over $\{0, \ldots, 2n\}$ such that, for all $t \in \mathbb{N}$ it holds that,

(a) if $X_t = x \notin \{0, 2n\}$, then $\Pr[X_{t+1} = x - 1] = \Pr[X_{t+1} = x + 1] = \frac{1}{2}$, and,
(b) if $X_t = x \in \{0, 2n\}$, then $\Pr[X_{t+1} = x] = 1$.

Further, let $T := \inf\{t \in \mathbb{N} \mid X_t \in \{0, 2n\}\}$.

Note that, for all $t \in \mathbb{N}$, it holds that $\mathrm{E}[X_{t+1} \mid X_t] = X_t$. Hence, we use Theorem 25 and bound, for all $t < T$,

$$\mathrm{Var}[X_{t+1} \mid X_t] = \frac{1}{2} \cdot (X_t - 1 - X_t)^2 + \frac{1}{2} \cdot (X_t + 1 - X_t)^2 = 1 .$$

Applying both Corollaries 26 and 27 yields the well-known result of

$$\mathrm{E}[T \mid X_0] = X_0(2n - X_0) .$$

Especially, for $X_0 = n$, we get $\mathrm{E}[T] = n^2$.

Declaration of competing interest

Conflict of interest does not apply.

References

[1] J. He, X. Yao, Drift analysis and average time complexity of evolutionary algorithms, Artif. Intell. 127 (2001) 57–85, https://doi.org/10.1016/S0004-3702(01)00058-3.
[2] J. He, X. Yao, A study of drift analysis for estimating computation time of evolutionary algorithms, Nat. Comput. 3 (2004) 21–35, https://doi.org/10.1023/B:NACO.0000023417.31393.c7.
[3] B. Hajek, Hitting-time and occupation-time bounds implied by drift analysis with applications, Adv. Appl. Probab. 14 (1982) 502–525.
[4] J. Lengler, Drift analysis, CoRR abs/1712.00964, arXiv:1712.00964, 2017.
[5] D. Johannsen, Random Combinatorial Structures and Randomized Search Heuristics, Ph.D. thesis, Universität des Saarlandes, 2010, http://scidok.sulb.uni-saarland.de/volltexte/2011/3529/pdf/Dissertation 3166 Joha Dani 2010.pdf.
[6] B. Doerr, D. Johannsen, C. Winzen, Multiplicative drift analysis, Algorithmica 64 (2012) 673–697, https://doi.org/10.1007/s00453-012-9622-x.
[7] P.K. Lehre, C. Witt, Concentrated Hitting Times of Randomized Search Heuristics with Variable Drift, Springer International Publishing, 2014, pp. 686–697.
[8] M.A. Semenov, D.A. Terkel, Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function, Evol. Comput. 11 (2003) 363–379, https://doi.org/10.1162/106365603322519279.
[9] D. Corus, D.-C. Dang, A.V. Eremeev, P.K. Lehre, Level-based analysis of genetic algorithms and other search processes, CoRR abs/1407.7663, http://arxiv.org/abs/1407.7663, 2014.
[10] P.K. Lehre, C. Witt, General drift analysis with tail bounds, CoRR abs/1307.2559, http://arxiv.org/abs/1307.2559, 2013.
[11] J. Lengler, A. Steger, Drift analysis and evolutionary algorithms revisited, Comb. Probab. Comput. 27 (2018) 643–666, https://doi.org/10.1017/S0963548318000275.
[12] B. Doerr, L.A. Goldberg, Adaptive drift analysis, Algorithmica 65 (2013) 224–250, https://doi.org/10.1007/s00453-011-9585-3.
[13] T. Kötzing, Concentration of first hitting times under additive drift, Algorithmica 75 (2016) 490–506, https://doi.org/10.1007/s00453-015-0048-0.
[14] P.S. Oliveto, C. Witt, Simplified drift analysis for proving lower bounds in evolutionary computation, Algorithmica 59 (2011) 369–386, https://doi.org/10.1007/s00453-010-9387-z.
[15] T. Kötzing, M.S. Krejca, First-hitting times under additive drift, in: Proc. of PPSN XV, 2018, pp. 92–104.
[16] G.R. Grimmett, D.R. Stirzaker, Probability and Random Processes, Oxford University Press, 2001.
[17] R. Bhattacharya, E.C. Waymire, A Basic Course in Probability Theory, Springer International Publishing, 2016.
[18] R. Durrett, Probability: Theory and Examples, Cambridge University Press, 2019.
[19] K. Azuma, Weighted sums of certain dependent random variables, Tohoku Math. J. 19 (1967) 357–367, https://doi.org/10.2748/tmj/1178243286.
[20] B. Mitavskiy, J.E. Rowe, C. Cannings, Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links, Int. J. Intell. Comput. Cybern. 2 (2009) 243–284, https://doi.org/10.1108/17563780910959893.
[21] J.E. Rowe, D. Sudholt, The choice of the offspring population size in the (1, λ) evolutionary algorithm, Theor. Comput. Sci. 545 (2014) 20–38, https://doi.org/10.1016/j.tcs.2013.09.036.
[22] B. Doerr, M. Fouz, C. Witt, Sharp bounds by probability-generating functions and variable drift, in: Proc. of GECCO'11, 2011, pp. 2083–2090.
[23] C. Gießen, C. Witt, Optimal mutation rates for the (1+λ) EA on onemax, in: Proc. of GECCO'16, 2016, pp. 1147–1154.
[24] B. Doerr, T. Kötzing, J.A.G. Lagodzinski, J. Lengler, Bounding bloat in genetic programming, in: Proc. of GECCO'17, 2017, pp. 921–928.
[25] D. Williams, Probability with Martingales, Cambridge University Press, 1991.