Algorithmica (2016) 75:490–506
DOI 10.1007/s00453-015-0048-0

Concentration of First Hitting Times Under Additive Drift

Timo Kötzing[1]

Received: 19 October 2014 / Accepted: 4 August 2015 / Published online: 11 August 2015

Abstract Recent advances in drift analysis have given us better and better tools for understanding random processes, including the run time of randomized search heuristics. In the setting of multiplicative drift we do not only have excellent bounds on the *expected* run time, but also more general results showing the *strong concentration* of the run time. In this paper we investigate the setting of additive drift under the assumption of strong concentration of the "step size" of the process. Under sufficiently strong drift towards the goal we show a strong concentration of the hitting time. In contrast to this, we show that in the presence of small drift a Gambler's-Ruin-like behavior of the process overrides the influence of the drift, leading to a maximal movement of about $\sqrt{t}$ steps within t iterations. Finally, in the presence of sufficiently strong negative drift the hitting time is superpolynomial with high probability; this corresponds to the well-known Negative Drift Theorem.

Keywords Additive drift · Concentration · Run time analysis

1 Introduction

Suppose we make a random walk on a finite set of real numbers starting at 0, stopping when we first hit some fixed value n. Further suppose that, in each step of the walk, we expect to increase in value by exactly $\varepsilon > 0$; this expected increase is called a positive *drift*. The Additive Drift Theorem ([11], see Theorem 5) tells us that the *expected* time for the walk to reach n for the first time is exactly n/ε. The (random) time to reach a given value for the first time is called the *first hitting time* or just

✉ Timo Kötzing
Timo.Koetzing@gmail.com

1 Fiedrich-Schiller-Universität Jena, 07743 Jena, Germany

hitting time; overshooting the given value will also be considered "hitting" in this paper. The Additive Drift Theorem is based on a more general result [10] and gave a new and powerful tool for the formal analysis of random optimization processes, such as the progress of randomized search heuristics (like evolutionary algorithms and ant colony optimization). For many randomized search heuristics such drift theorems are particularly useful as the algorithms can be described as making a (biased) random walk through the search space. In order to bound the time until reaching a certain part of the search space (for example a global optimum), one typically derives bounds on the expected progress per iteration; these bounds can then be turned into an expected time until reaching the desired part of the search space by using a drift theorem.

After the publication of He and Yao [11] the Additive Drift Theorem became more and more popular as a method to analyze the expected run time of randomized search heuristics. In order to get better bounds from a drift theorem with little effort, new drift theorems were proven, for example for drift proportional to the distance from the target (instead of uniform drift, as in the Additive Drift Theorem—this is called *multiplicative drift*) [5]. Another very powerful family of drift theorems are the so-called *Variable Drift Theorems* (independently developed in [12, Theorem 4.6] and [16, Section 8], but see also [21] for a discussion and extension).

All these theorems have in common that they can be used for showing *upper* bounds on the run time of randomized algorithms. Aiming for a similarly strong tool for showing lower bounds [17,18], derived (again from [10]) a theorem which applies in case that the drift goes *away* from the target (see Theorem 6 for a precise statement and [21] for a powerful variant). Just as the drift theorems for upper bounds, the Negative Drift Theorem has proven to be tremendously useful for the analysis of randomized search heuristics, providing an easy-to-apply tool for deriving lower bounds.

In addition to bounds on the *expected* hitting time, concentration results are also of interest. These can, for example, be directly used for statements about the concentration of the run time of an algorithm. But sometimes concentration results are necessary for deriving bounds on the expected hitting times as well: imagine, for example, an algorithm which can only be successful when n independent sub-algorithms are successful; in the analysis, one would usually need concentration results for the run time of the sub-algorithms.

For the special case of multiplicative drift, strong concentration results were given in [4]. In very recent work [14] even more general results are given, providing concentration bounds in a very general setting. In this paper, we take an approach different from that in [14] by focusing on the very special case of additive drift and deriving as strong as possible concentration results in this case. The advantage is that, for the theorems in this paper, checking whether they apply is easy, and so is using the conclusion; the downside is the restricted scope.

Outside of the evolutionary computation community, a number of results also regarding positive drift are known. In particular, most of the work of this paper is based on the technique of bounded differences (basically all proofs are applications of the Azuma–Hoeffding Inequality), which is wide-spread, and the applications given here are straightforward instances of these methods (see [7] for an introduction). A notable example where similar results regarding positive drift are shown is [22] (a version of Theorem 1 can be found in its Section 4). The main purpose of the present paper is to introduce these methods to the evolutionary computation community in an easily accessible way (the Azuma–Hoeffding Inequality itself was already used in the community to derive concentration, see, for example [6]).

All results in this paper hold for the case where the step size of the random variables is bounded by some constant c. In order to extend the scope of the theorems we use the concept of *sub-Gaussian random variables* [2] and allow the step sizes to be any such random variable. Example sub-Gaussian random variables include bounded variables, but also exponentially decaying random variables (see Sect. 3.1). This way all theorems are now applicable to a wide range of processes such as the ones occurring in the analysis of randomized search heuristics. We refer the reader to Fan et al. [9] to an excellent collection of many useful bounds in the context of drift.

This paper is an extended version of Kötzing [13]; in particular, this earlier version does not cover all sub-Gaussian random variables, but only those bounded by a constant.

1.1 Discussion of Results

Recall that, if we start at 0 and drift an expected amount of at most ε towards our goal $n > 0$, we have an expected time of at least n/ε to reach n. However, it is possible that n is already reached after one round with constant probability: the process might, in the first iteration, jump to the goal ($= n$) with probability $1/2$ and with the remaining $1/2$ probability it jumps to $-n$, giving an expected progress of $0 \leq \varepsilon$.[1] Similarly, one can give examples where the drift is high, but the probability to reach the goal within the expected number of steps is low.

We would like to give sufficient conditions (which hold in many cases for analyses of randomized search heuristics) under which the hitting time is concentrated around the expectation. To that end we will assume that large jumps are very unlikely. Formally, we will require that the progress in each iteration is a random variable Δ such that, for some constants c and δ,

$$\forall z \in [0, \delta] : \mathrm{E}(\exp(z\Delta)) \leq \exp(z^2 c/2). \tag{1}$$

We call any random variable Δ which fulfills Eq. (1) a (c, δ)-*sub-Gaussian random variable* ([2]; see also [3] for a discussion on the concept of sub-Gaussian random variables). Intuitively this concept captures that there are no large jumps: any proof of the Azuma-Hoeffding Inequality uses that any random variable on $[-c, c]$ is (c^2, ∞)-sub-Gaussian; furthermore, in Theorem 10 we see that any exponentially decaying random variable is also sub-Gaussian (with parameters that depend on the speed of decay).

Under the condition that the progress in each iteration is sub-Gaussian we can derive strong results; Fig. 1 gives an overview. Note the three regimes of additive drift: if it

[1] Note that iterating this idea leads to an example where, under arbitrary additive drift, the expected number of iterations until n is reached is 2, seemingly contradicting the Additive Drift Theorem; however, this iterated example requires an unbounded search space, which is ruled out by the requirements of the Additive Drift Theorem.

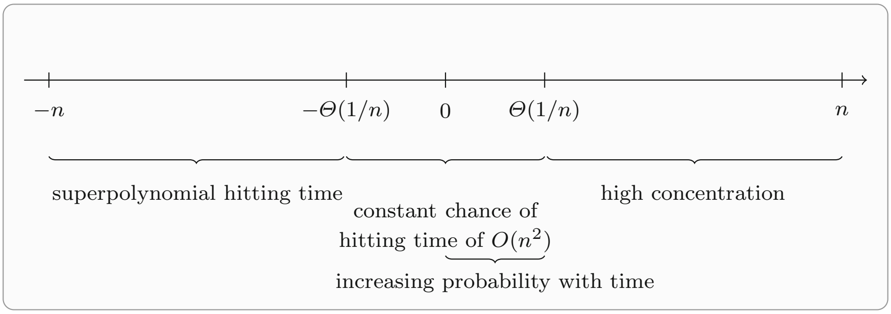

Fig. 1 Intuitive regimes of additive drift. Depicted are possible values of the additive drift ε for constant step sizes; the important change points are (up to constant factors) at $1/n$ and $-1/n$. Note that for large (superconstant) bounds on the possible jump size c of the process, the results get worse

is strong, we get high concentration; if it is between about $-\Theta(1/n)$ and $\Theta(1/n)$, we get a behavior similarly to the Gambler's Ruin problem, with a constant probability of reaching n regardless of the strength of the drift due to a (sufficiently unbiased) random walk on the real line; note that this result requires constant variance. This constant probability can be significantly boosted by allowing more time, in case of non-negative drift. Finally, for strongly negative values of drift (this is the regime of Negative Drift Theorems), we get an exponential hitting time with superpolynomial probability. In the following we discuss these statements in more detail; for simplicity, we will only discuss the case of random processes with bounded step width; later we will generalize these statements to arbitrary sub-Gaussian steps.

Our first theorem informs about an exponentially small probability of arriving at n significantly before the expected n/ε iterations. Note that a version of this bound was shown in [22, Corollary 4.1].

Theorem 1 *Let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \leq 0\}$ we denote the random variable describing the earliest point that the random process exceeds n, given a starting value of at most 0. Suppose there are $\varepsilon, c > 0$ such that, for all t,*

1. $E(X_{t+1} - X_t \mid X_0, \dots X_t, T > t) \leq \varepsilon$ *(an additive drift of at most ε), and*
2. $|X_t - X_{t+1}| < c$ *(bounded step width).*

Then, for all $s \leq n/(2\varepsilon)$,

$$P(T < s) \leq \exp\left(-\frac{n^2}{8c^2 s}\right).$$

We see that, for example for constant c and $\varepsilon = O(1/n)$, we have a superpolynomially small probability of hitting n in less than $n^2/\omega(\log n)$ iterations. Note that the bound is no longer useful (i.e. greater than 1) when $s \geq n^2$. This means that after more than n^2 steps we cannot exclude having exceeded n (at least not with this theorem). If $\varepsilon \geq 1/n$, we expected to hit n after n^2 steps anyway (due to the drift), and the bound of $s \leq n/(2\varepsilon)$ makes the bound inapplicable for values of $s \geq n^2$. As soon as we have drift of $\varepsilon < 1/n$, the drift process is intuitively more and more drowned by the random walk due to the variance (which we will consider later).

But what is now the probability of arriving significantly after the expected time? For that we need a *lower* bound on the expected progress (drift).

Theorem 2 *Let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \geq 0\}$ we denote the random variable describing the earliest point that the random process exceeds n, given a starting value of at least 0. Suppose there are $\varepsilon, c > 0$ such that, for all t,*

1. $\mathrm{E}(X_{t+1} - X_t \mid X_0, \dots X_t, T > t) \geq \varepsilon$, *and*
2. $|X_t - X_{t+1}| < c$.

Then, for all $s \geq 2n/\varepsilon$,

$$\mathrm{P}(T \geq s) \leq \exp\left(-\frac{s\varepsilon^2}{8c^2}\right).$$

Thus, unless the drift is small, n will be exceeded with high probability after twice the expected number of steps. For small drift ($O(1/n)$), the bound is only meaningful for larger numbers of iterations, so that Markov's Inequality will give better bounds in this case for s close to n/ε.

If the drift is significantly negative, then we cannot hope to reach the goal in polynomial time with reasonable probability; this is the statement of the Negative Drift Theorem ([17,18], see Theorem 6). A scaled version can be found in [19, Theorem 22], which also implies the following theorem.

Theorem 3 ([19]) *Let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \leq 0\}$ we denote the random variable describing the earliest point that the random process exceeds n, given a starting value of at most 0. Suppose there are c, $0 < c < n$ and $\varepsilon < 0$ such that, for all t,*

1. $\mathrm{E}(X_{t+1} - X_t \mid X_0, \dots X_t, T > t) \leq \varepsilon$, *and*
2. $|X_t - X_{t+1}| < c$.

Then, for all $s \geq 0$,

$$\mathrm{P}(T \leq s) \leq s \exp\left(-\frac{n|\varepsilon|}{2c^2}\right).$$

For example, for a constant c and $\varepsilon = -\omega(\log(n)/n)$, this gives a superpolynomially small hitting probability for any polynomial number of steps.

Finally, we consider the case where there is only small drift $\varepsilon \in [0, 1/n]$.

Theorem 4 *Let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation and let $n > 0$. With $T = \min\{t \geq 0 : X_t \geq n \mid X_0 \geq 0\}$ we denote the random variable describing the earliest point that the random process exceeds n, given a starting value of at least 0. Suppose there is c with $0 < c < n$ such that, for all t,*

1. $\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots X_t, T > t) \geq 0$,
2. $\mathrm{Var}(X_{t+1} - X_t \mid X_0, \ldots X_t, T > t) \geq 1$, *and*
3. $|X_t - X_{t+1}| < c$.

Then there is a constant ℓ (independent of n, c and ε) such that, for all $p > 0$,

$$\mathrm{P}(T \leq n^2/p^{\ell \log(c)}) \geq 1 - p.$$

For example, if we have a constant c and want any constant hitting probability δ, then a quadratic number of steps suffices (just as in the Gambler's Ruin problem).

2 Known Bounds

The literature knows a large number of drift theorems; we give the two most important ones with respect to our setting of additive drift.

First we give the classic *Additive Drift Theorem*.

Theorem 5 (Additive Drift [11]) *Let $(X_t)_{t\geq 0}$ be random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}$. Let T be the random variable that denotes the earliest point in time $t \geq 0$ such that $X_t \geq n$. If there exists $\varepsilon > 0$ such that, for all $t > 0$,*

$$\mathrm{E}(X_{t+1} - X_t \mid T > t) \leq \varepsilon,$$

then

$$\mathrm{E}(T \mid X_0) \geq \frac{X_0}{\varepsilon}.$$

If there exists $\varepsilon > 0$ such that, for all t,

$$\mathrm{E}(X_{t+1} - X_t \mid T > t) \geq \varepsilon,$$

then

$$\mathrm{E}(T \mid X_0) \leq \frac{X_0}{\varepsilon}.$$

Second, the *Negative Drift Theorem* concerns adverse drift and shows a high hitting time, which can be used to derive lower bounds on the run time of algorithms.

Theorem 6 (Negative Drift [17,18]) *Let $(X_t)_{t\geq 0}$ be real-valued random variables describing a stochastic process over some state space. Suppose there is an interval $[a, b] \subseteq \mathbb{R}$, two constants $\delta, \varepsilon > 0$ and, possibly depending on $\ell = b - a$, a function $r(\ell)$ satisfying $1 \leq r(\ell) = o(\ell/\log \ell)$ such that, for all $t \geq 0$, the following conditions hold.*

1. $\mathrm{E}(X_{t+1} - X_t \mid a < X_t < b) \geq \varepsilon$;
2. *For all* $j \geq 0$, $P(|X_{t+1} - X_t| \geq j \mid a < X_t) \leq \frac{r(\ell)}{(1+\delta)^j}$.

Then there is a constant c such that, for $T = \min\{t \geq 0 : X_t \leq a \mid X_0 \geq b\}$, *we have*

$$\mathrm{P}(T \leq 2^{c\ell/r(\ell)}) = 2^{-\Omega(\ell/r(\ell))}.$$

A crucial requirement of the theorem is a restriction on the jump size of the random process: the larger the step, the less likely it must be. A further important requirement is that of *constant* drift away from the target; this requirement can be circumvented via scaling, see [19, Theorem 22]. See Corollary 22 for a comparison with the results of this paper.

A sequence of random variables $(X_t)_{t\geq 0}$ is called a *supermartingale* if each random variable has finite expectation and, for all $t \geq 0$,

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots X_t) \leq 0.$$

How is additive drift related to the concept of (super)martingales? In the presence of additive drift of at most ε, we have, for all $t \geq 0$, the inequality

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots X_t) \leq \varepsilon.$$

This means that $(X_t - t\varepsilon)_{t\geq 0}$ *is* a supermartingale, making all the strong and well-developed machinery for martingales applicable. For our results, we make use of a variant of the Azuma–Hoeffding Inequality for supermartingales, see [1]. We give a version from [8, Corollary 2.1] for reference, but we will cite a more flexible version in Sect. 3.2.

Theorem 7 (Azuma–Hoeffding Inequality) *Let* $(X_t)_{t\geq 0}$ *be a supermartingale such that there is a sequence* $(c_t)_{t\geq 0}$ *of real number such that, for all* $t \geq 0$, $|X_{t+1} - X_t| < c_t$. *For all t let* $C_t = \sum_{i=0}^{t-1} c_t^2$. *Then, for all* $t \geq 0$ *and all* $x > 0$,

$$\mathrm{P}\left(\max_{0\leq j\leq t}(X_j - X_0) \geq x\right) \leq \exp\left(-\frac{x^2}{2C_t}\right).$$

The Azuma–Hoeffding Inequality is sometimes stated with $X_t - X_0$ in place of $\max_{0\leq j\leq t}(X_j - X_0)$. However, as discussed at the end of Section 3.5 in [15], the stronger version typically comes at no extra price.

3 Bounds for Sub-Gaussian Random Variables

In this section we discuss sub-Gaussian random variables and give a variant of the Azuma-Hoeffding Inequality. In particular, we want to extend the Azuma-Hoeffding Inequality to sequences of random variables $(X_t)_{t\geq 0}$ where the differences $X_{t+1} - X_t$

are not bounded, but *sub-Gaussian* as follows. Formally, we call a sequences of random variables $(X_t)_{t\geq 0}$ $((c_t)_{t\geq 0}, \delta)$-*sub-Gaussian* iff, for all $t \geq 0$,

$$\forall z \in [0, \delta] : \mathrm{E}(\exp(z(X_{t+1} - X_t)) \mid X_0, \ldots, X_t) \leq \exp(z^2 c_t/2). \tag{2}$$

We allow for $\delta = \infty$ with the obvious meaning.[2] In case that there is a c such that, for all j, $c_j = c$, we simplify notation and call $(X_t)_{t\geq 0}$ a (c, δ)-sub-Gaussian.

The following remark follows from calculus (see, for example, [20] for an exposition).[3] In particular, any sub-Gaussian sequences of random variables is always a submartingale.

Remark 8 Let $(X_t)_{t\geq 0}$ be $((c_t)_{t\geq 0}, \delta)$-sub-Gaussian. Then, for all $t \geq 0$,

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots, X_t) \leq 0$$

and

$$\mathrm{Var}(X_{t+1} - X_t \mid X_0, \ldots, X_t) \leq c_t.$$

We proceed by giving examples for sub-Gaussian supermartingales, before giving bounds for these random processes.

3.1 Examples for Sub-Gaussian Supermartingales

We start with the simple example of a bounded supermartingale.

Theorem 9 *Let $(X_t)_{t\geq 0}$ be a supermartingale such that there is a $c > 0$ with*

$$\forall t \geq 0 : |X_{t+1} - X_t| \leq c.$$

Then $(X_t)_{t\geq 0}$ is (c^2, ∞)-sub-Gaussian.

The straightforward proof of this fact is at the heart of the proof of the Azuma-Hoeffding Inequality as given in Theorem 7. Next we see that also exponentially decaying random variables are sub-Gaussian.

Theorem 10 *Let $(X_t)_{t\geq 0}$ be a supermartingale such that there are $c > 0$ and δ with $0 < \delta < 1$ and, for all $t \geq 0$,*

$$\forall x \geq 0 : \mathrm{P}(|X_{t+1} - X_t| \geq x \mid X_0, \ldots, X_t) \leq \frac{c}{(1+\delta)^x}. \tag{3}$$

Then $(X_t)_{t\geq 0}$ is $(128c\delta^{-3}, \delta/4)$-sub-Gaussian.

[2] In fact, the term *sub-Gaussian* originally entailed $\delta = \infty$, while the version with finite δ is called *locally sub-Gaussian* [3]. Furthermore, this terminology is usually applied to single random variables, not to martingales.

[3] Note that Eq. (2) is typically required to also hold for negative $z \geq -\delta$, in which case even stronger statements can be made. However, we want to keep the scope of these definitions wide.

Proof We will need the following equation from calculus.

$$\forall a > 0 : \int_0^\infty x^2 e^{-ax} dx = \frac{2}{a^3}. \tag{4}$$

Let f be the probability density function of $|X_{t+1} - X_t|$ for given $X_0, \ldots, X_t$ and let z be such that $0 \leq z \leq \ln(1+\delta)/2$; note that from $\delta < 1$ we know $\ln(1+\delta) \geq \delta/2$, so that we indeed consider all z with $0 \leq z \leq \delta/4$ as necessary. It is easy to see that, for all real numbers x, $e^x \leq 1 + x + x^2 e^{|x|}/2$. Therefore, we have

$$\begin{aligned}
&\mathrm{E}(\exp(z(X_{t+1} - X_t)) \mid X_0, \ldots, X_t)\\
&\quad\leq 1 + z\,\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots, X_t)\\
&\qquad + \frac{z^2}{2}\,\mathrm{E}((X_{t+1} - X_t)^2 \exp(z|X_{t+1} - X_t|) \mid X_0, \ldots, X_t)\\
&\quad\leq 1 + \frac{z^2}{2}\,\mathrm{E}((X_{t+1} - X_t)^2 \exp(z|X_{t+1} - X_t|) \mid X_0, \ldots, X_t)\\
&\quad\leq 1 + \frac{z^2}{2}\int_0^\infty x^2 e^{zx}\,\mathrm{P}(|X_{t+1} - X_t| \geq x \mid X_0, \ldots, X_t)\,\mathrm{d}x\\
&\quad= 1 + \frac{z^2}{2}\int_0^\infty x^2 f(x) e^{zx}\,\mathrm{d}x.
\end{aligned}$$

Note that in the second inequality we use $\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots, X_t) \leq 0$ (the supermartingale property). From Eq. (3) we have

$$\forall x : f(x) \leq \frac{c}{(1+\delta)^x}.$$

In order to abbreviate terms in the next chain of inequalities, we let

$$B = \frac{2c}{(\ln(1+\delta) - z)^3}.$$

Thus, we can extend the above chain of equalities and inequalities with

$$\begin{aligned}
&\mathrm{E}(\exp(z(X_{t+1} - X_t)) \mid X_0, \ldots, X_t)\\
&\quad\leq 1 + \frac{z^2}{2}\int_0^\infty x^2 \frac{c}{(1+\delta)^x} e^{zx}\,\mathrm{d}x\\
&\quad= 1 + \frac{cz^2}{2}\int_0^\infty x^2 e^{-x(\ln(1+\delta)-z)}\,\mathrm{d}x\\
&\quad= 1 + \frac{cz^2}{2}\frac{2}{(\ln(1+\delta) - z)^3}\\
&\quad= 1 + z^2 B/2\\
&\quad\leq e^{z^2 B/2}.
\end{aligned}$$

For $z \leq \ln(1+\delta)/2$ and using $\ln(1+\delta) \geq \delta/2$ we know

$$B \leq \frac{2c}{(\ln(1+\delta)/2)^3} = \frac{16c}{\ln(1+\delta)^3} \leq \frac{128c}{\delta^3}.$$

This shows the desired result.

3.2 Bounds for Sub-Gaussian Supermartingales

Now we turn to using sub-Gaussian supermartingales to derive bounds for first hitting times. The following is a version of the Azuma–Hoeffing Inequality modified to sub-Gaussian supermartingales taken from [9].

Theorem 11 ([9]) *Let $(X_t)_{t\geq 0}$ be $((c_t)_{t\geq 0}, \delta)$-sub-Gaussian. For all $t \geq 0$, let*

$$C_t = \sum_{j=0}^{t-1} c_j.$$

Then, for all $t \geq 0$ and all $x > 0$,

$$\mathrm{P}\left(\max_{0\leq j\leq t}(X_j - X_0) \geq x\right) \leq \exp\left(-\frac{x}{2}\min\left(\delta, \frac{x}{C_t}\right)\right).$$

Proof The statement of this theorem follows from [9, Theorem 2.6] by using the following parameters. We let $n = t$; for all $i \geq 1$, let V_{i-1} be the random variable which is constantly c_i; let $v^2 = C_t$; for all λ, let $f(\lambda) = \lambda^2/2$. Now we choose

$$\lambda = \min\left(\delta, \frac{x}{C_t}\right)$$

and get the result directly from [9, Theorem 2.6] (after a few small manipulations).

The following is now a straightforward corollary.

Theorem 12 *Let $(X_t)_{t\geq 0}$ be a sequence of random variables and let $d \in \mathbb{R}$. If, for all $t \geq 0$,*

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots, X_t) \leq d,$$

then $(X_t - dt)_{t\geq 0}$ is a supermartingale. If further $(X_t - dt)_{t\geq 0}$ is (c, δ)-sub-Gaussian, then, for all $t \geq 0$ and all $x > 0$,

$$\mathrm{P}\left(\max_{0\leq j\leq t}(X_j - X_0) \geq dt + x\right) \leq \exp\left(-\frac{x}{2}\min\left(\delta, \frac{x}{ct}\right)\right).$$

The following is a corollary which regards the probability of undershooting.

Theorem 13 *Let $(X_t)_{t\geq 0}$ be a sequence of random variables. If there is d such that, for all $t \geq 0$,*

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots, X_t) \geq d,$$

then $(dt - X_t)_{t\geq 0}$ is a supermartingale. If further $(dt - X_t)_{t\geq 0}$ is (c, δ)-sub-Gaussian, then, for all $t \geq 0$ and all $x > 0$,

$$\mathrm{P}\left(\max_{0\leq j\leq t}(X_j - X_0) \leq dt - x\right) \leq \exp\left(-\frac{x}{2}\min\left(\delta, \frac{x}{ct}\right)\right).$$

4 Detailed Theorems and Proofs

In this section we generalize the theorems from Sect. 1 to sub-Gaussian supermartingales. The proofs are applications of the (generalized) Azuma-Hoeffding Inequality and its corollaries from Sect. 3.2; we will discuss these in Sect. 4.1 on large drift (they mostly apply when the drift is $\Omega(1/n)$ in either direction). After that we will consider small drift in Sect. 4.2.

For this section, let $(X_t)_{t\geq 0}$ be random variables over $\mathbb{R}$, each with finite expectation. Furthermore, we let $n \in \mathbb{N}$ and let $T_{\leq} = \min\{t \geq 0 : X_t \geq n \mid X_0 \leq 0\}$ be the random variable that denotes hitting time of n (similarly, $T_{\geq} = \min\{t \geq 0 : X_t \geq n \mid X_0 \geq 0\}$). For a given $\varepsilon > 0$, we say that $(X_t)_{t\geq 0}$ *has drift of at most* ε iff, for all $t \geq 0$,

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots X_t, T > t) \leq \varepsilon. \tag{5}$$

Symmetrically, we say, for a given $\varepsilon > 0$, that $(X_t)_{t\geq 0}$ *has drift of at least* ε iff, for all $t \geq 0$,

$$\mathrm{E}(X_{t+1} - X_t \mid X_0, \ldots X_t, T > t) \geq \varepsilon. \tag{6}$$

4.1 Large Drift

We will now use the (generalized) Azuma–Hoeffding Inequality to extend Theorems 1–3.

Theorem 14 (Extending Theorem 1) *Suppose that $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon > 0$ and $(X_t - \varepsilon t)_{t\geq 0}$ is (c, δ)-sub-Gaussian. Then, for all $s \leq n/(2\varepsilon)$,*

$$\mathrm{P}(T_{\leq} < s) \leq \exp\left(-\frac{n}{4}\min\left(\delta, \frac{n}{2cs}\right)\right).$$

Proof Let $s \leq n/(2\varepsilon)$. We apply Theorem 12 (with $x = n/2$ and $t = s$). Intuitively, we bound the probability of gaining twice the distance that we should have gained.

Similarly, we get a bound showing a high hitting probability after sufficiently many iterations.

Theorem 15 (Extending Theorem 2) *Suppose that* $(X_t)_{t\geq 0}$ *has drift of at least* $\varepsilon > 0$ *and* $(\varepsilon t - X_t)_{t\geq 0}$ *is* (c, δ)*-sub-Gaussian. Then, for all* $s \geq 2n/\varepsilon$,

$$\mathrm{P}(T_{\geq} \geq s) \leq \exp\left(-\frac{s\varepsilon}{4}\min\left(\delta, \frac{\varepsilon}{2c}\right)\right).$$

Proof Let $s \geq 2n/\varepsilon$. We apply Theorem 13 (with $x = s\varepsilon/2$ and $t = s$). Intuitively, we bound the probability of gaining only half the distance that we should have gained; this is meaningful once we should have overshot by a factor of 2, i.e. for $s \geq 2n/\varepsilon$ as desired.

We now use the same approach to prove the theorem concerning negative drift.

Theorem 16 (Extending Theorem 3) *Suppose that* $(X_t)_{t\geq 0}$ *has drift of at most* $\varepsilon < 0$ *and* $(X_t - \varepsilon t)_{t\geq 0}$ *is* (c, δ)*-sub-Gaussian. Then, for all* $s \geq 0$,

$$\mathrm{P}(T_{\leq} \leq s) \leq s\exp\left(-\frac{n}{2}\min\left(\delta, \frac{|\varepsilon|}{c}\right)\right).$$

Proof We make an analysis with phases. A phase begins at t if $X_t < 0$ and $X_{t+1} \geq 0$ and ends at t' if either $X_{t'} \geq n$ or $X_{t'} < 0$; in the first case we call the phase *successful*, in the second case *unsuccessful*. We will show that a phase is successful with probability at most $\exp(-n/2\min(\delta, |\varepsilon|/c))$, as then a union bound (or an application of Bernoulli's Inequality) will give the desired result, lower bounding the length of each phase with the trivial bound of 1. In order to bound the probability of a phase being successful, we use the following reasoning. Any phase starts ≤ 0. If the process does not overshoot its expectation by n ever within $n/|\varepsilon|$ iterations, it not only did not reach n (starting from ≤ 0) but is certainly below 0 (as, after $n/|\varepsilon|$ iterations, the expectation is $\leq -n$). To bound the probability of this event we apply again Theorem 12 (with $x = n$ and $t = n/|\varepsilon|$) to see that the probability of a phase being successful is at most

$$\exp\left(-\frac{n}{2}\min\left(\delta, \frac{|\varepsilon|}{c}\right)\right),$$

as desired.

4.2 Small Drift

We start with a lemma which is interesting in its own right, showing the theorem concerning negative drift (Theorem 3) to be reasonably tight. The proof of the lemma makes use of one of the results concerning the concentration of the hitting time under positive drift (Theorem 1).

Lemma 17 *Let* $b = 3072$*. Then, for all* n *and* $c > 0$*, the following holds. Let* $k = bcn$*. Suppose that* $(X_t)_{t\geq 0}$ *has a drift of at least* $\varepsilon \geq -1/(8k)$ *and a variance in each step of at least* 1*. Also suppose* $(X_t - \varepsilon t)_{t\geq 0}$ *is* (c, δ)*-sub-Gaussian, with* $\delta = \omega(1/n)$*. Let*

$s = 24bcn^2$. *Then we have that, within s steps, the process does not drop below* $-k$ *with probability* $\geq 1/2$ *and*

$$\mathrm{P}\left(T_{\geq} \leq s\right) \geq \frac{1}{2}.$$

Proof We give the proof for $\varepsilon \leq 0$; the case of $\varepsilon > 0$ is analogous, but easier. We let A be the event that the process does not drop below $-k$ within s steps. We first show $P(A) \geq 3/4$, after that we show that, conditional on A, the process reaches n with probability 3/4, which will imply the claim.

For all $t \geq 0$, we let

$$Y_t = (X_t)^2$$

and

$$\Delta_t = X_{t+1} - X_t.$$

We can assume, without loss of generality, $\mathrm{E}(\Delta_t) \leq 0$. In all of the following computations of expectation and variance the conditioning on all relevant (previous) random variables is implicitly understood but not made explicit for clarity (and brevity) of the exposition. From Remark 8 we know that, for all $t \geq 0$, $\mathrm{Var}(\Delta_t) \leq c$ (which also implies $c \geq 1$, from our lower bound on the variance). It suffices to show that Y_t does not reach k^2 within s steps with probability $\geq 3/4$. We want to apply Theorem 14 to $(Y_t)_{t\geq 0}$, so we compute the expected drift.

$$\begin{aligned}
\mathrm{E}(Y_{t+1}) &= \mathrm{E}((\Delta_t + X_t)^2)\\
&= \mathrm{E}((\Delta_t)^2 + 2\Delta_t X_t + X_t^2)\\
&= \mathrm{E}((\Delta_t)^2) + 2\,\mathrm{E}(\Delta_t)X_t + X_t^2\\
&= \mathrm{Var}(\Delta_t) + E(\Delta_t)^2 + 2\,\mathrm{E}(\Delta_t)X_t + Y_t\\
&\quad \leq c + 1 + Y_t\\
&\quad \leq 2c + Y_t,
\end{aligned}$$

where the first inequality follows from our bound on the variance, together with $-1/(8k) \leq \mathrm{E}(\Delta_t) \leq 0$ and $X_t \geq -k$.

In order to estimate the number of steps until $(X_t)_{t\geq 0}$ reaches $-k$, we wait until the process drops below 0, and bound the time that the process $(Y_t)_{t\geq 0}$ takes to get from 0 to k^2, which, using the Additive Drift Theorem, has an expectation of at least $k^2/(2c) \geq 2s$ steps. As $(Y_{t+1}-Y_t) = (X_{t+1}-X_t)(X_{t+1}+X_t)$ and $X_{t+1}+X_t \leq 2n$, we get that the process $(Y_t - 2ct)_{t\geq 0}$ is $(4n^2c, \delta/(2n))$-sub-Gaussian. Thus, Theorem 14 gives that $(Y_t)_{t\geq 0}$ *does* exceed k^2 within s steps starting from 0 with probability at most

$$\exp\left(-\frac{k^4}{4\cdot 8n^2cs}\right) = \exp\left(-\frac{b^3c^2}{32\cdot 24}\right) \leq 1/4$$

as desired.

Now we want to bound the probability for reaching n. To this end we let, for all $t \geq 0$,

$$Z_t = (X_t + k)^2 - k^2$$

and we condition on A. In a computation analogous to that for $(Y_t)_{t\geq 0}$ we see that

$$\mathrm{E}(Z_{t+1}) \geq 1/2 + Z_t.$$

From the Additive Drift Theorem (Theorem 5) we now know that $(Z_t)_{t\geq 0}$ reaches $(n+k)^2 - k^2$ in an *expected* number of at most

$$2((n+k)^2 - k^2) = 2(2nk + n^2) \leq 6nk$$

steps. Thus, using Markov's Inequality, we get that $(Z_t)_{t\geq 0}$ reaches $(n+k)^2 - k^2$ within s steps with probability at least $3/4$ as desired, as $s = 4(6nk)$.

Next we use this lemma to get a direct corollary, basically extending the idea of Gambler's Ruin.

Theorem 18 *Suppose that* $(X_t)_{t\geq 0}$ *has a drift of at least* $\varepsilon = -o(cn)$ *and a variance in each step of at least* 1*. Also suppose* $(X_t - \varepsilon t)_{t\geq 0}$ *is* (c, δ)*-sub-Gaussian, with* $\delta = \omega(1/n)$*. Then, for some* $s = O(cn^2)$*,*

$$\mathrm{P}\left(T_{\geq} \leq s\right) \geq \frac{1}{2}.$$

Now we want to boost this probability of $1/2$ arbitrarily high by allowing longer run times. The idea is to apply Lemma 17 iteratively and get arbitrarily good bound with an induction.

Theorem 19 (Extending Theorem 4) *Let* $b = 3072$ *just as in Lemma* 17*. Suppose that* $(X_t)_{t\geq 0}$ *has drift of at least* $\varepsilon \geq 0$*, a variance in each step of at least* 1 *and* $(X_t - \varepsilon t)_{t\geq 0}$ *is* (c, δ)*-sub-Gaussian, with* $\delta = \omega(1/n)$*. Then there is a constant* ℓ *(independent of* n*,* c *and* ε*) such that, for all* $p > 0$*,*

$$\mathrm{P}(T_{\geq} \leq n^2/p^{\ell \log(c)}) \geq 1 - p.$$

Proof We let $k_0 = 0$ and, for all i,

$$k_{i+1} = bc(k_i + n) + n + k_i$$

and

$$s_i = 24bc(k_i + n)^2.$$

We analyze the process in an infinite series of phases, starting with phase 0. For each i, Phase $i+1$ starts as soon as Phase i ends (Phase 0 starts at time $t=0$). Phase i ends when either the goal is reached (the process is $\geq n$), the process is at most $-k_{i+1}$, or s_i steps passed in Phase i, whichever happens first. We call a phase *successful* if it ends with reaching the goal.

Trivially, just before the beginning of Phase i, the process is at least $-k_i$. We want to apply Lemma 17, where the n of the Lemma corresponds to k_i+n (for the application of the Lemma, we shift the process by k_i). Thus, we see that each phase is successful with probability at least $1/2$.

Let $p>0$ and let $a>0$ be such that $2^{a-1}\geq 1/p\geq 2^a$. Thus, after a phases, we have a success probability of at least $1-p$ as desired. As, for all i, Phase i takes at most s_i steps, we get the desired result.

5 Corollaries

In this section we will derive some useful corollaries to our theorems. We will use the terminology of the preceding section.

The first corollary is derived from Theorems 14 and 15 and gives an interval in which the hitting time is with high probability. Note that the interval is smaller if ε is larger than $\Theta(1/n)$.

Corollary 20 (Concentration) *Let $(X_t)_{t\geq 0}$ with $X_0=0$ and let T be the hitting time of n; let $\varepsilon=\Omega(1/n)$ be given. Suppose there are constants y, y' such that $(X_t)_{t\geq 0}$ has a drift of at most $y\varepsilon$ and at least $y'\varepsilon$. Furthermore, suppose that $(X_t-y\varepsilon t)_{t\geq 0}$ and $(y'\varepsilon t-X_t)_{t\geq 0}$ are (c,δ)-sub-Gaussian where $\delta=\omega(\log(n)/n)$. Then, for each k, there is a k' (independent of n and ε) such that*

$$\mathrm{P}\left((n/\varepsilon)/(k'c\log n)\leq T\leq k'(n/\varepsilon)c\log n\right)\geq 1-n^{-k}.$$

Furthermore, for all $r=\omega(c\log n)$,

$$\mathrm{P}\left((n/\varepsilon)/r\leq T\leq (n/\varepsilon)r\right)\geq 1-n^{-\omega(1)}.$$

The next corollary is derived from Theorem 16. It shows that sufficiently negative drift gives strong (lower) bounds on the hitting time.

Corollary 21 (Negative Drift) *Suppose that $(X_t)_{t\geq 0}$ has drift of at most $\varepsilon = -\omega(c\log n/n)$ and assume $(X_t-\varepsilon t)_{t\geq 0}$ is (c,δ)-sub-Gaussian with $\delta=\omega(\log(n)/n)$. Then, for all polynomials p and all n large enough,*

$$\mathrm{P}(T_{\leq}\leq p(n))\leq \frac{1}{p(n)}.$$

We can also recover the Negative Drift Theorem ([17,18], see Theorem 6) from Theorem 16 (more precisely, as a corollary to its proof, where it is easy to see that negative drift is only required in some bounded interval). Note that, in order to follow the notation of Theorem 6, the process attempts to go *down* (from b to a), so what is called a "negative drift" is a positive value (away from the goal). This also requires an application of Theorem 10 to get the result for exponentially decaying step width. If instead we try to get a similar statement where we use the exponential decay to establish that the process has a bounded step width (with sufficiently high probability), we can get the following corollary.

Corollary 22 (Negative Drift II) *Suppose there is an interval* $[a, b] \subseteq \mathbb{R}$, *two constants* $\delta, \varepsilon > 0$ *and, possibly depending on* $\ell = b - a$, *a function* $r(\ell)$ *satisfying* $1 \leq r(\ell) = \exp(o(\sqrt[4]{\ell}))$ *such that, for all* $t \geq 0$, *the following conditions hold.*

1. $\mathrm{E}(X_{t+1} - X_t \mid a < X_t < b) \geq \varepsilon$;
2. *For all* $j \geq 0$, $\mathrm{P}(|X_{t+1} - X_t - \varepsilon| \geq j \mid a < X_t) \leq \frac{r(\ell)}{(1+\delta)^j}$.

Then there is a constant c such that, for $T = \min\{t \geq 0 : X_t \leq a \mid X_0 \geq b\}$, *we have*

$$\mathrm{P}\left(T \leq 2^{c\sqrt{\ell}}\right) = 2^{-\Omega\left(\sqrt[4]{\ell}\right)}.$$

This corollary can give good bounds where the version of the Negative Drift Theorem given in Theorem 6 is not applicable; this is for example the case for $r(\ell) = \ell$. Also in some cases where both are applicable the corollary gives slightly better bounds: Consider for example the case of $r(\ell) = \ell/(\log n)^2$. Corollary 22 gives superpolynomial run time with just the same probability as for smaller r, while Theorem 6 gives

$$\mathrm{P}(T \leq 2^{c(\log \ell)^2}) = 2^{-\Omega((\log \ell)^2)}.$$

Note that this is also a superpolynomial run time with superpolynomially high probability.

Acknowledgments I would like to thank Tobias Friedrich, Benjamin Doerr and Anton Krohmer for many useful discussions on the topic of this paper; all of them also provided valuable pointers to the literature. A further source of pointers to the literature as well as many helpful comments came from the reviewers of the conference version this paper, who have proved to be very knowledgeable in this subject. Most importantly, Carsten Witt pointed me to possible extensions of the Azuma–Hoeffding Inequality, as well as to some computations in his own publications, which helped to extend this paper to cases of non-bounded variables, making the contribution significantly more valuable; this paper would have a much more restricted scope without his advice. Finally, I would like to thank the reviewers of the journal version of this paper; they pointed out places for correction and improvement, as well as more useful literature.

References

1. Azuma, K.: Weighted sums of certain dependent random variables. Tohoku Math. J. **19**, 357–367 (1967)
2. Buldygin, V.V., Kozachenko, Y.V.: Sub-Gaussian random variables. Ukr. Math. J. **32**, 483–489 (1980)
3. Chareka, P., Chareka, O., Kennedy, S.: Locally sub-Gaussian random variables and the strong law of large numbers. Atl. Electron. J. Math. **1**, 75–81 (2006)
4. Doerr, B., Goldberg, L.A.: Adaptive drift analysis. Algorithmica **65**, 224–250 (2013)
5. Doerr, B., Johannsen, D., Winzen, C.: Multiplicative drift analysis. Algorithmica **64**, 673–697 (2012)
6. Doerr, B., Künnemann, M.: How the (1+λ) evolutionary algorithm optimizes linear functions. In: Proceedings of GECCO (Genetic and Evolutionary Computation Conference), pp. 1589–1596 (2013)

7. Dubhashi, D.P., Panconesi, A.: Concentration of Measure for the Analysis of Randomized Algorithms. Cambridge University Press, Cambridge (2009)
8. Fan, X., Grama, I., Liu, Q.: Hoeffding's inequality for supermartingales. Stoch. Process. Appl. **122**, 3545–3559 (2012)
9. Fan, X., Grama, I., Liu, Q.: Exponential inequalities for martingales with applications. Electron. J. Probab. **20**, 1–22 (2015)
10. Hajek, B.: Hitting-time and occupation-time bounds implied by drift analysis with applications. Adv. Appl. Probab. **13**, 502–525 (1982)
11. He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Nat. Comput. **3**, 21–35 (2004)
12. Johannsen, D.: Random Combinatorial Structures and Randomized Search Heuristics. Ph.D. thesis, Universität des Saarlandes. Available online at http://scidok.sulb.uni-saarland.de/volltexte/2011/3529/pdf/Dissertation_3166_Joha_Dani_2010 (2010)
13. Kötzing, T.: Concentration of first hitting times under additive drift. In: Proceedings of GECCO (Genetic and Evolutionary Computation Conference)
14. Lehre, P.K., Witt, C.: Concentrated hitting times of randomized search heuristics with variable drift. In: Proceedings of ISAAC (International Symposium on Algorithms and Computation), pp. 686–697 (2014)
15. McDiarmid, C.: Concentration. In: Habib, M., McDiarmid, C., Ramirez-Alfonsin, J., Reed, B. (eds.) Probabilistic Methods for Algorithmic Discrete Mathematics. Algorithms and Combinatorics, Vol. 16, pp. 195–248. Springer, Berlin, Heidelberg (1998)
16. Mitavskiy, B., Rowe, J.E., Cannings, C.: Preliminary theoretical analysis of a local search algorithm to optimize network communication subject to preserving the total number of links. In: Proceedings of CEC (Congress on Evolutionary Computation), pp. 1484–1491 (2008)
17. Oliveto, P.S., Witt, C.: Simplified drift analysis for proving lower bounds in evolutionary computation. Algorithmica **59**, 369–386 (2011)
18. Oliveto, P.S., Witt, C.: Erratum: simplified drift analysis for proving lower bounds in evolutionary computation. 1211.7184 (2012)
19. Oliveto, P.S., Witt, C.: On the runtime analysis of the simple genetic algorithm. Theor. Comput. Sci. **545**, 2–19 (2014)
20. Rivasplata, O.: Subgaussian random variables: an expository note. http://www.math.ualberta.ca/~orivasplata/publications/subgaussians (2012)
21. Rowe, J.E., Sudholt, D.: The choice of the offspring population size in the (1, λ) EA. In: Proceedings of GECCO (Genetic and Evolutionary Computation Conference), pp. 1349–1356 (2012)
22. Wormald, N.C.: The differential equation method for random graph processes and greedy algorithms. In: Lectures on Approximation and Randomized Algorithms, pp. 73–155 (1999)