Additive drift with tail bounds

Vihnin F. Antipov D. Sinyachenko N.

October 1, 2021

1 Introduction

Drift analysis is a powerful toolbox for the analysis of random processes. Taking the roots from Hajek's negative drift [Haj82] drift analysis was significantly developed by the evolutionary computation community starting with a seminal work by He and Yao [HY04] introducing the additive drift theorem. Since then many modifications have been produced such as multiplicative drift [DJW12] and variable drift [Joh10, RS12]. The most recent overview of the family of drift theorems can be found in [Len17].

All drift theorems consider a random process $\{X_t\}_{t \in \mathbb{N}}$, which starts with some $X_0 = a$ and aims to reach some value b. **TODO: Adapt all theorems to this notation or change the notation, if it is easier.** Without loss of generality, we assume that $a < b$[1]. The *hitting time* T for such process is the minimum t such that $X_t \geq b$. The main idea of drift analysis is to transform the available information about the process such as the expected step size $E[X_{t+1} - X_t \mid X_t = s]$ (which is called the *drift* of the process) into information about the hitting time. E.g., the additive drift theorem claims that if the expected step size is bounded from below by some $\delta > 0$, then the expected hitting time is at most the distance $(b - a)$ which the process has to cover divided by δ (with some additional conditions such as X_t cannot jump over its goal, that is, $X_t \leq b$ for all $t \in \mathbb{N}$). More formally, we have

$$E[T] \leq \frac{b-a}{\delta}.$$

The additive drift theorem also works into an opposite direction. That is, if the drift is bounded from above by some $\delta > 0$, then we have

$$E[T] \geq \frac{b-a}{\delta}.$$

Several other drift theorems are capable to give us more information about the hitting time. In particular, the multiplicative drift theorem [DJW12] gives an exponentially small

[1] Some of the drift theorems are formulated for a process which starts at a greater value b and aims to reach a smaller value a. However, such results can be inverted by considering a process $Y_t = -X_t$.

bound on the probability of the hitting time to exceed its expectation. The reason why the additive drift theorem does not give us any similar bounds on the concentration of the hitting time is that it uses only a small piece of information about the considered process, which is, the expected progress. To be able to obtain some tail bounds on the hitting time we should also have some bounds on the concentration of the steps of the process. For example, consider two processes $\{X_t\}_{t \in \mathbb{N}}$ and $\{Y_t\}_{t \in \mathbb{N}}$ which start with $X_0 = Y_0 = 0$ and the hitting time for these processes is when they reach 100. Let $\{X_t\}_{t \in \mathbb{N}}$ be a deterministic process such that for all $t \in N$ we have $X_{t+1} = X_t + 1$. Let $\{Y_t\}_{t \in \mathbb{N}}$ be a process such that with probability $\frac{1}{100}$ we have $Y_{t+1} = Y_t + 100$ and we have $Y_{t+1} = Y_t$ otherwise. Both processes have the same drift equal to one, and both processes reach 100 in the same expected time, namely in 100 steps. However, the hitting time of $\{X_t\}_{t \in \mathbb{N}}$ is determined and has zero variation, while the hitting time of $\{Y_t\}_{t \in \mathbb{N}}$ follows a geometric distribution $\text{Geom}(\frac{1}{100})$ and its standard deviation is $10\sqrt{99} \approx 99$.

In practice we often meet something in the middle between these two processes, namely processes with non-deterministic, but well-concentrated steps. Intuition suggests that for such processes the variation of the hitting time should not be large too. We are aware of two approaches for the analysis of such processes, which are, the analysis via sub-Gaussian processes shown in [Köt16] and the analysis via the negative drift theorem shown in [ADK19]. We discuss both approaches and show that they have the same nature, however, they have different conditions when they are applicable. **TODO: State exact results here: the differences and similarities.**

2 Existing tools

The two approaches for deriving the tail bounds on the hitting time for the processes under additive drift shown in [Köt16] and in [ADK19] use the same trick. They decompose the random process X_t into a sum of a deterministic process and a super- or sub-martingale. Namely, they consider

$$X_t = Y_t + \varepsilon t,$$

where ε is a constant chosen in such a way that Y_t is either za super- or sub-martingale.

When we aim at showing the lower tail bound, that is, for some time bound B we want to estimate $\Pr[T \leq B]$, then we must

1. know the *upper* bound δ on the expected progress of X_t and
2. choose ε which is at least δ so that Y_t was a super-martingale.

After that the main goal is to show that Y_t does not exceed some value $c > 0$ in at least B steps, since it would imply that X_t does not exceed $\varepsilon t + c$. Hence, if we consider only B such that $(b - a) > \varepsilon B + c$ then we have

$$\Pr[T \leq B] = \Pr[\forall t \in [1..B]\ X_t < b] = \Pr[\forall t \in [1..B]\ Y_t < c].$$

Note that this works only for $B < \frac{b-a-c}{\varepsilon}$, which is less than the expected hitting time, since by the additive drift theorem we have

$$E[T] \geq \frac{b-a}{\delta} > \frac{b-a-c}{\varepsilon}.$$

The main difference between the methods shown in [Köt16] and [ADK19] is in how they estimate $\Pr[\forall t \in [1..B]\ Y_t < c]$. In [Köt16] we consider Y_t which are sub-Gaussian processes and estimate the probability that they do not exceed c using some tools specific for the sub-Gaussian processes. In [ADK19] we consider Y_t which are a subject to the Hajek's negative drift theorem.

We describe these two approaches in more details in this section.

We also note that we can obtain the upper tail bounds (that is, estimates on the probability $\Pr[T > B]$) with similar arguments. For this we also need to use a decomposition $X_t = Y_t + \varepsilon t$, but now we require

1. the expected progress to be bounded from below by some $\delta > 0$ and
2. Y_t to be a sub-martingale.

2.1 Negative drift

In this section we discuss the Hajek's negative drift theorem, which is most commonly used in the following form.

Theorem 2.1 (Drift Theorem for Lower Bounds)**.** *Let $\{X_t\}_{t \geq 0}$ be a Markow process over a finite set of states $\mathbb{S}$, and $\mathbf{g} : \mathbb{S} \to \mathbb{R}$ a function that assigns to every state a non-negative real numbers. Pick two real numbers a and b such that $a < b$ and let the random variable T denote the earliest point in time $t \geq 0$ where $\mathbf{g}(X_t) \leq a$ holds.*

If there are constants $\lambda > 0$, $D \geq 1$, and a $p > 1$ taking only positive values, for which the following four conditions hold

(1) $\mathbf{g}(X_0) \geq b$,

(2) $\forall t \geq 0 E\left[e^{-\lambda(\mathbf{g}(X_{t+1} - \mathbf{g}(X_t)))} \mid X_t,\ \mathbf{g}(X_t) < b\right] \leq 1 - \frac{1}{p} =: \rho$,

(3) $\forall t \geq 0 E\left[e^{-\lambda(\mathbf{g}(X_{t+1} - b))} \mid X_t,\ \mathbf{g}(X_t) \geq b\right] \leq D$,

then for all time bounds $B \geq 0$ the probability that T exceeds B is at most

$$\Pr[T \leq B] \leq e^{\lambda(a-b)} \cdot B \cdot D \cdot p \tag{1}$$

This form of the theorem is not stated in [Haj82] and we are not aware of any proof which derives it from the results of Hajek's paper. Hence we present a simple proof of the negative drift theorem which is based on the following lemma from [Haj82].

Lemma 2.2 (Lemma 2.8 in [?]). *If conditions 2 and 3 are met then for all $t \geq 0$ we have*

$$\Pr[\mathbf{g}(X_t) \leq a] \leq \rho^t e^{\lambda(a-\mathbf{g}(X_0))} + \frac{1-\rho^t}{1-\rho} D e^{\lambda(a-b)}$$

Proof of Theorem 2.1. If $T = B$, then we have $\mathbf{g}(X_k) \leq a$ and for all $t < B$ we have $\mathbf{g}(X_t) > a$. Hence, we compute

$$\begin{aligned}
\Pr[T \leq B] &= \sum_{k=1}^{B} \Pr[T = k] \\
&= \sum_{k=1}^{B} \Pr[\mathbf{g}(X_k) \leq a \wedge \mathbf{g}(X_{k-1}) > a \wedge ... \wedge \mathbf{g}(X_0) > a)] \\
&\leq \sum_{k=1}^{B} \Pr[\mathbf{g}(X_k) \leq a].
\end{aligned}$$

By Lemma 2.2

$$\begin{aligned}
\Pr[T \leq B] &\leq \sum_{k=1}^{B} \left(\rho^k e^{\lambda(a-\mathbf{g}(X_0))} + \frac{1-\rho^k}{1-\rho} D e^{\lambda(a-b)} \right) \\
&= \rho \left(\frac{1-\rho^B}{1-\rho} \right) e^{\lambda a} \left(e^{-\lambda \mathbf{g}(X_0)} - \frac{D}{1-\rho} e^{-\lambda b} \right) + e^{\lambda(a-b)} \frac{BD}{1-\rho}.
\end{aligned}$$

Since $\mathbf{g}(X_0) \geq b$ we have

$$\Pr[T \leq B] \leq \rho \frac{1-\rho^B}{1-\rho} e^{\lambda(a-b)} \left(1 - \frac{D}{1-\rho} \right) + e^{\lambda(a-b)} \frac{BD}{1-\rho}$$

Since $\frac{D}{1-\rho} \geq D \geq 1$, we further compute

$$\Pr[T \leq B] \leq e^{\lambda(a-b)} \frac{BD}{1-\rho} = e^{\lambda(a-b)} \cdot B \cdot D \cdot p$$

□

2.2 Theorem for sub-gaussian processes

coming soon...

2.3 Comparing requirements

The theorem for sub-Gaussian processes is applicable, when there exist $\delta \geq 0$ and $c \in \mathbb{R}$ such that for all $t \geq 0$ and for all $\gamma \in [0, \delta]$ we have

$$E\left[e^{\gamma(X_{t+1}-X_t)}\right] \leq e^{\frac{c\gamma^2}{2}}.$$

At the same time the negative drift theorem is applicable when there exist $\delta \geq 0$ and $p \geq 1$ such that for all $t \geq 0$ we have

$$E\left[e^{\gamma(X_{t+1}-X_t)}\right] \leq 1 - \frac{1}{p} \, .$$

To compare rigors of requirements we need to explore behavior $E[e^{\gamma X}]$ depending to γ. This behavior depends on the distribution X, hence we consider the two cases for continuos and for discrete distributions.

Continuos distribution

By the definition of expectation of a function we have

$$E[e^{\gamma X}] = \int_{-\infty}^{+\infty} f(x) e^{\gamma x} \, dx,$$

where $f(x)$ is a probability density function for X.

We compute the first derivative of this expectation over γ

$$\frac{\partial E[e^{\gamma X}]}{\partial \gamma} = \frac{\partial}{\partial \gamma} \int_{-\infty}^{+\infty} f(x) e^{\gamma x} \, dx = \int_{-\infty}^{+\infty} f(x) \frac{\partial e^{\gamma x}}{\partial \gamma} \, dx = \int_{-\infty}^{+\infty} x f(x) e^{\gamma x} \, dx =$$
$$= \int_{-\infty}^{0} x f(x) e^{\gamma x} \, dx + \int_{0}^{+\infty} x f(x) e^{\gamma x} \, dx.$$

Note that for all $x \leq 0$ we have $e^{\gamma x} \leq 1$ and for all $x \geq 0$ we have $e^{\gamma x} \geq 1 + \gamma x$. Hence since $f(x) \geq 0$ for all x, we have

$$\int_{0}^{+\infty} x f(x) e^{\gamma x} \, dx \geq \int_{0}^{+\infty} x f(x)(1 + \gamma x) \, dx$$
$$\int_{-\infty}^{0} x f(x) e^{\gamma x} \, dx \geq \int_{-\infty}^{0} x f(x) \, dx.$$

Therefore,

$$\begin{aligned}\frac{\partial E[e^{\gamma X}]}{\partial \gamma} &\geq \int_0^{+\infty} x f(x)(1+\gamma x)\, dx + \int_{-\infty}^{0} x f(x)\, dx \\ &= E[X] + \gamma \int_0^{+\infty} x^2 f(x)\, dx.\end{aligned}$$

We denote $c = \int_0^{+\infty} x^2 f(x)\, dx \geq 0$.If we integrate the inequality above, we obtain a lower bound on the expectation.

$$E[e^{\gamma X}] = E[1] + \int_0^{\gamma} \frac{\partial E[e^{\gamma X}]}{\partial \gamma}\, d\gamma \geq 1 + \gamma E[X] + \gamma^2 \frac{c}{2}.$$

Discrete distribution

In the same way we consider the expectation of function by definition

$$E[e^{\gamma X}] = \sum_{i=1}^{+\infty} e^{\gamma x_i} \Pr[x = x_i].$$

Denote set $P = \{i \in \mathbb{N} : x_i \geq 0\}$. Hence

$$E[e^{\gamma X}] = \sum_{i=1}^{+\infty} e^{\gamma x_i} Pr[x = x_i] = \sum_{i \in P} e^{\gamma x_i} Pr[x = x_i] + \sum_{i \in \mathbb{N} \setminus P} e^{\gamma x_i} Pr[x = x_i],$$

When we take the first derivative we can split to two group, first of which are terms with indices from P set and others

$$\begin{aligned}\frac{\partial E[e^{\gamma X}]}{\partial \gamma} &= \sum_{i=1}^{+\infty} \frac{\partial e^{\gamma x_i}}{\partial \gamma} Pr[x = x_i] = \sum_{i=1}^{+\infty} x_i e^{\gamma x_i} Pr[x = x_i] \\ &= \sum_{i \in P} x_i e^{\gamma x_i} Pr[x = x_i] + \sum_{i \in \mathbb{N} \setminus P} x_i e^{\gamma x_i} Pr[x = x_i] \\ &\geq \sum_{i \in P} x_i (1 + \gamma x_i) Pr[x = x_i] + \sum_{i \in \mathbb{N} \setminus P} x_i Pr[x = x_i] \\ &= E[X] + \gamma \sum_{i \in P} x_i^2 Pr[x = x_i]\end{aligned}$$

We denote $c = \sum_{i \in P} x_i^2 Pr[x = x_i] \geq 0$ and obtain a bound similar to the one for conditions random variables.

$$E[e^{\gamma X}] \geq 1 + \gamma E[X] + \gamma^2 \frac{c}{2}$$

Since the bounds for both cases are the same, we do not distinguish cases for continuos and discrete v.v.

2.3.1 Impracticability for positive drift

In this case, the expectation of the exponent cannot be lower than 1, and also in the some neighborhood of zero it grows faster than any $e^{\frac{\gamma^2 c}{2}}$, since

$$\frac{\partial e^{\frac{\gamma^2 c}{2}}}{\partial \gamma} = c\gamma e^{\frac{\gamma^2 c}{2}} = g(\gamma),$$

Note that $g(0) = 0$, while $E[X] > 0$, therefore, that in any neighborhood of zero we have

$$\forall c > 0\, \exists \gamma_c > 0\, \forall \gamma \in (0, \gamma_c) e^{\frac{\gamma^2 c}{2}} < E[e^{\gamma X}],$$

Hence, both requirements are not feasible.

2.3.2 Zero drfit case

In this section we consider some specific family of continuos distributions with symmetric probability density functions, because their expectation equals 0.

Exponent of polynomial

Consider function

$$f(x) = ce^{-|x|^a},\ a > 0.$$

Compute c

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{-\infty}^{+\infty} ce^{-|x|^a}\, dx = 1,$$

$$\int_{-\infty}^{+\infty} e^{-|x|^a}\, dx = 2\int_{0}^{+\infty} e^{-x^a}\, dx.$$

We denote

$$\begin{aligned} u &= x^a \\ du &= dx\,(ax^{a-1}) = dx\,(au^{1-\frac{1}{a}}), \end{aligned}$$

Then, we have

$$2\int_{0}^{+\infty} e^{-x^a}\, dx = \frac{2}{a}\int_{0}^{+\infty} e^{-u} u^{\frac{1}{a}-1}\, dx = \frac{2\Gamma(\frac{1}{a})}{a}.$$

We conclude that $c = \frac{a}{2\Gamma(\frac{1}{a})}$ and

$$f(x) = \frac{a}{2\Gamma(\frac{1}{a})} e^{-|x|^a}.$$

We compute the n-th momentum of X (expectation от X^n) as follows

$$E[X^n] = \int_{-\infty}^{+\infty} f(x)x^n \, dx = \frac{\Gamma(\frac{n+1}{a})}{2\Gamma(\frac{1}{a})}(1 + (-1)^n)$$

Then we can consider $e^{\gamma X}$ as a infinite series

$$E[e^{\gamma X}] = E\left[\sum_{n=0}^{+\infty} \frac{(\gamma X)^n}{n!}\right] = \sum_{n=0}^{+\infty} \gamma^n \frac{E[X^n]}{n!} = \sum_{n=0}^{+\infty} \gamma^{2n} \frac{E[X^{2n}]}{2n!} =$$
$$= \sum_{n=0}^{+\infty} \frac{(\gamma^2)^n}{n!} \left(\frac{E[X^{2n}]n!}{2n!}\right) = \sum_{n=0}^{+\infty} \frac{(\gamma^2)^n}{n!} a_n,$$

Where by a_n we denote $\frac{E[X^{2n}]n!}{2n!}$ for all $n \in \mathbb{N}$.

Third equality is satisfied because all odd momentums are 0 due to the symmetry of the function.

Consider the limit of the sequence $\{\sqrt[n]{a_n}\}_{n\in\mathbb{N}}$

$$A = \lim_{n\to+\infty} \sqrt[n]{a_n} = \lim_{n\to+\infty} \sqrt[n]{\frac{n!E[X^{2n}]}{2n!}} = \lim_{n\to+\infty} \sqrt[n]{\frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\, \Gamma\left(\frac{2n+1}{a}\right)}{\left(\frac{2n}{e}\right)^{2n} \sqrt{4\pi n}\, \Gamma\left(\frac{1}{a}\right)}}$$
$$= \lim_{n\to+\infty} \frac{n}{e}\frac{e^2}{4n^2} \sqrt[n]{\Gamma\left(\frac{2n+1}{a}\right)} \simeq \lim_{n\to+\infty} \frac{e}{4n} \sqrt[n]{\left(\frac{2n+1}{ae}\right)^{\frac{2n+1}{a}}}$$
$$= \lim_{n\to+\infty} \frac{e}{4n}\left(\frac{2n}{ae}\right)^{\frac{2}{a}} = \left(\frac{e^{1-\frac{2}{a}}}{2^{2-\frac{2}{a}} a^{\frac{2}{a}}}\right) \lim_{n\to+\infty} n^{\frac{2}{a}-1}.$$

We now consider two cases depending on the value of a

Case 1: $a \geq 2$

Then A exist, which implies that there exist $C > 0$ such that $\forall n \geq 0 a_n \leq C^n$. Hence $E[e^{\gamma X}] \leq e^{\gamma^2 C}$.

Case 2: $a < 2$

Sequence doesn't converge, so you can't say anything about sub-gausality.

We can compute same limit for $\{\sqrt[n]{\frac{a_n}{n!}}\}_{n\in\mathbb{N}}$ and threshold regarding to a is 1, where only for $a \geq 1$ series converge.

In this case we need to explore behavior of series when $a \in (1, 2)$. By another definition of sub-gaussian process

$$\exists K > 0\ \forall p \geq 1\, (E[X^p])^{\frac{1}{p}} \leq K\sqrt{p}.$$

Let's compute (assume $p \vdots 2$)

$$E[X^p] = \frac{\Gamma\left(\frac{p+1}{a}\right)}{\Gamma\left(\frac{1}{a}\right)}$$

$$E[X^p]^{\frac{1}{p}} \underset{p\to+\infty}{=} \left(\frac{p}{ea}\right)^{\frac{p}{a}\frac{1}{p}} = \mathcal{O}(p^{\frac{1}{a}}).$$

For $a \in (1, 2)$ the asymptotic is greater that square root. Hence sub-gausality is not feasible.

2.3.3 Processes with negative drift

In this case, similarly to the positive drift case, it is the summand equal to the expectation that makes the main contribution to the derivative in some neighborhood of zero, since

$$\frac{\partial E[e^{\gamma x}]}{\partial \gamma}\Big|_{\gamma=0} = E[X] < 0.$$

Hence both requirements are feasible, since we have

$$\exists \gamma_0 > 0 \forall \gamma \in (0, \gamma_0) E[e^{\gamma X}] < 1.$$

Consequently proving that the process is sub-gaussian is tantamount to finding p in the second requirements.

РАССУЖДЕНИЯ ПРО П И КАРТИНКИ

Define function for distribution X B_X such that

$$B_X(\gamma) = 1 + \gamma E[X] + \frac{\gamma^2}{2} E[X^2]_+,$$

Where

$$E[X^2]_+ = \begin{cases} \int_0^{+\infty} x^2 f(x)\, dx, & \text{if } X \text{ is continuos} \\ \sum_{x\geq 0} x^2 \Pr[X = x], & \text{otherwise.} \end{cases}$$

Also define function $Ee_X(\gamma) = E[e^{\gamma X}]$. We proofed in end of section 2.3 that

$$\forall \gamma \geq 0 : Ee_X(\gamma) \geq B_X(\gamma).$$

If $Ee(\gamma)$ is a differentiable function, then by definition the derivative will be continuos. Since $e^{\gamma x}$ and $f(x)$ is non-negative for all x and γ, hence

$$\frac{\partial^2 Ee(\gamma)}{\partial \gamma^2} = \int_{-\infty}^{+\infty} x^2 f(x) e^{\gamma x}\, dx > 0.$$

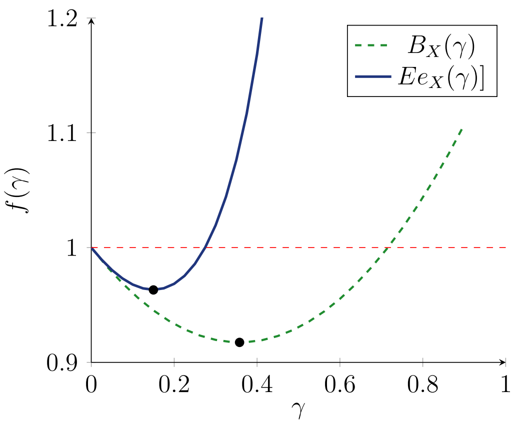
Then

$$\left.\frac{\partial Ee(\gamma)}{\partial \gamma}\right|_{\gamma=\gamma_0} - \left.\frac{\partial Ee(\gamma)}{\partial \gamma}\right|_{\gamma=0} = \int_0^{\gamma_0} \frac{\partial^2 Ee(\gamma)}{\partial \gamma^2}\, d\gamma > 0$$
$$\frac{\partial Ee\,(\gamma_0)}{\partial \gamma} = E[X] + \int_0^{\gamma_0} \frac{\partial^2 Ee(\gamma)}{\partial \gamma^2}\, d\gamma$$

Since second derivative always great than zero, than first derivative is continuously growing.

Then if $E[X] < 0$ $Ee(\gamma)$ would go down in some neighborhood. And if $E[X^2]_+ \neq 0$ it will have point with zero derivative, what fits the condition of the second requirement.

Example for distribution X with probability density function $f(x) = 0.65e^{-0.65(x+2)}$:

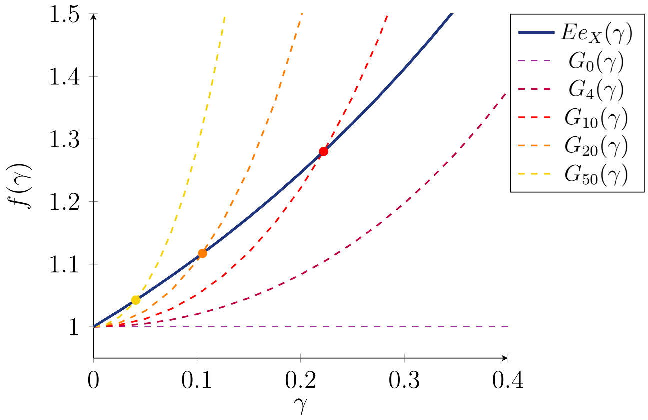


If $E[X] > 0$ then in some neighborhood around $\gamma = 0$ would be greater than any $e^{\frac{c\gamma^2}{2}}$. To proof it, consider function $G_c(\gamma) = e^{\frac{c\gamma^2}{2}}$, then

$$\frac{\partial G_c(\gamma)}{\partial \gamma} = c\gamma\, G_c(\gamma) = g(\gamma),$$

$g(0) = 0 < E[X]$ and $Ee_X(0) = B_X(0) = 1$. Hence both of requirements are not feasible.

Example for $X \sim \mathcal{N}(1,1)$:



Conclusions

In the case of **zero** expectation we cannot always guarantee sub-gausality, which has been proved by the example of series of even functions $e^{-|x|^a}$.

In the case of **positive** expectation both criteria are not feasible, and if it is **negative**, then both are feasible.

3 Tail bounds

3.1 Upper bounds

Consider a process $\{X_t\}_{t\in\mathbb{N}}$ with positive drift (i.e. $E[X_{t+1} - X_t|\ X_t = s] \geq 0$ for all s) and another process $\{Y_t\}_{t\in\mathbb{N}}$ such that $Y_t = X_t - \varepsilon t$ and it has negative drift (i.e. $E[Y_{t+1} - Y_t|\ Y_t = s] \leq 0$ for all s).

Our aim is to bound the probability $\Pr[T_X \leq t_0]$, where T_X is the first time when $X_t \geq b$ for some $b > X_0$ and for t_0. We first note that

$$\Pr[T_X \leq t_0] \leq \Pr[X_{t_0} \geq b] \leq \Pr[Y_{t_0} \geq b - \varepsilon t_0].$$

Since $\{Y_t\}_{t\in\mathbb{N}}$ is a process with negative drift, it is a subject to Theorem []. Hence, we have

$$\Pr[Y_{t_0} \geq b - \varepsilon t_0] \leq t_0 D p e^{-\gamma(b-\varepsilon t_0 - a)},$$

which also implies

$$\Pr[T_X \le t_0] \le t_0 D p e^{-\gamma(b-\varepsilon t_0 - a)}.$$

Let $t_0 := \frac{\kappa(b-a)}{\varepsilon}$. Then we compute

$$\Pr\left[T_X \le \frac{\kappa(b-a)}{\varepsilon}\right] \le \frac{\kappa(b-a)}{\varepsilon} D p e^{-\gamma(1-\kappa)(b-a)}.$$

For example we can use $\kappa = \frac{1}{2}$ and obtain

$$\Pr\left[T_X \le \frac{(b-a)}{2\varepsilon}\right] \le \frac{(b-a)}{2\varepsilon} D p e^{-\gamma \frac{(b-a)}{2}}.$$

3.2 Lower bounds

Consider a process $\{X_t\}_{t\in\mathbb{N}}$ with positive drift (i.e. $E[X_{t+1} - X_t|\ X_t = s] \ge 0$ for all s) and another process $\{Y_t\}_{t\in\mathbb{N}}$ such that $Y_t = \varepsilon t - X_t$ and it has negative drift (i.e. $E[Y_{t+1} - Y_t|\ Y_t = s] \le 0$ for all s).

Our aim is to bound the probability $\Pr[T_X > t_0]$, where T_X is the first time when $X_t \ge b$ for some $b > X_0$ and for t_0. We first note that

$$\Pr[T_X > t_0] \le \Pr[X_{t_0} < b] \le \Pr[Y_{t_0} > \varepsilon t_0 - b].$$

Since $\{Y_t\}_{t\in\mathbb{N}}$ is a process with negative drift, it is a subject to Theorem []. Hence, we have

$$\Pr[Y_t > \varepsilon t_0 - b] \le t_0 D p e^{-\gamma(\varepsilon t_0 - b + a)},$$

which also implies

$$\Pr[T_X > t_0] \le t_0 D p e^{-\gamma(\varepsilon t_0 - b + a)}.$$

Let $t_0 := \frac{\kappa(b-a)}{\varepsilon}$. Then we compute

$$\Pr\left[T_X > \frac{\kappa(b-a)}{\varepsilon}\right] \le \frac{\kappa(b-a)}{\varepsilon} D p e^{-\gamma(\kappa-1)(b-a)}.$$

For example we can use $\kappa = \frac{3}{2}$ and obtain

$$\Pr\left[T_X > \frac{3(b-a)}{2\varepsilon}\right] \le \frac{3(b-a)}{2\varepsilon} D p e^{-\gamma \frac{(b-a)}{2}}.$$

4 Time bounds for process with variable λ

coming soon...

4.1 Case when λ is a random variable

Consider the total progress ΔX made in T iterations. It is a sum of all progresses ΔX_i made in each of T iterations.

$$\Delta X = \sum_{i=0}^{T} \Delta X_i$$

By Wald's equality we have

$$E[\Delta X] = E\left[\sum_{i=1}^{T} \Delta X_i\right].$$

If we assume that $E[\Delta X_i]$ is a constant or at least we can give relatively sharp lower and upper bounds on it, then

$$E[\Delta X] = E\left[\sum_{i=1}^{T} \Delta X_i\right] \approx E[T]E[\Delta X_i].$$

Let Λ be the total cost, which is the total number of fitness evaluations (the fitness evaluations in iteration i equals λ_i). Since λ is a random variable, each λ_i has the same expectation. Therefore, we have

$$E\left[\Lambda\right] = E\left[\sum_{i=1}^{T} \lambda_i\right] = E[T]E[\lambda] \approx \frac{E[\Delta X]E[\lambda]}{E[\Delta X_i]}$$

If we denote $\vartheta_\lambda = \frac{E[\Delta X_i]}{E[\lambda]}$, which is the expected progress per fitness evaluation, then the previous equation is simplified to

$$E[\Lambda] \approx \frac{E[\Delta X]}{\vartheta_\lambda}.$$

5 Conclusion

coming soon...

References

[ADK19] Denis Antipov, Benjamin Doerr, and Vitalii Karavaev. A tight runtime analysis for the $(1 + (\lambda, \lambda))$ GA on LeadingOnes. In *Foundations of Genetic Algorithms, FOGA 2019*, pages 169–182. ACM, 2019.

[DJW12] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Multiplicative drift analysis. *Algorithmica*, 64:673–697, 2012.

[Haj82] Bruce Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14:502–525, 1982.

[HY04] Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3:21–35, 2004.

[Joh10] Daniel Johannsen. *Random Combinatorial Structures and Randomized Search Heuristics*. PhD thesis, Universität des Saarlandes, 2010.

[Köt16] Timo Kötzing. Concentration of first hitting times under additive drift. *Algorithmica*, 75:490–506, 2016.

[Len17] Johannes Lengler. Drift analysis. *CoRR*, abs/1712.00964, 2017.

[RS12] Jonathan E. Rowe and Dirk Sudholt. The choice of the offspring population size in the (1,λ) EA. In *Proc. of GECCO'12*, pages 1349–1356. ACM, 2012.