Additive drift with tail bounds

Vihnin F. Antipov D. Sinyachenko N.

September 26, 2021

1 Introduction

coming soon...

2 Existing tools

coming soon...

2.1 Negative drift

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

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

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

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

To prove this theorem we use the following lemma from [**?**]

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

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

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

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

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

$$\begin{aligned}\int_{0}^{+\infty} xf(x)e^{\gamma x}\,dx &\geq \int_{0}^{+\infty} xf(x)(1+\gamma x)\,dx \\ \int_{-\infty}^{0} xf(x)e^{\gamma x}\,dx &\geq \int_{-\infty}^{0} xf(x)\,dx.\end{aligned}$$

Therefore,

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

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

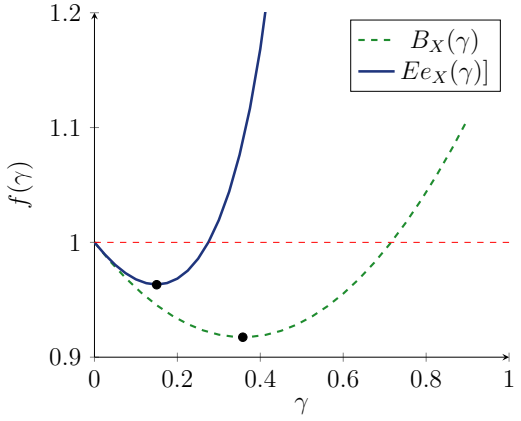
$$\frac{\partial Ee(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_0} - \frac{\partial Ee(\gamma)}{\partial \gamma}\Big|_{\gamma=0} = \int_0^{\gamma_0} \frac{\partial^2 Ee(\gamma)}{\partial \gamma^2}\, d\gamma > 0$$

$$\frac{\partial Ee\,(\gamma_0)}{\partial \gamma} = E[X] + \int_0^{\gamma_0} \frac{\partial^2 Ee(\gamma)}{\partial \gamma^2}\, d\gamma$$

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 $f(x) = 0.65e^{-0.65(x+2)}$:

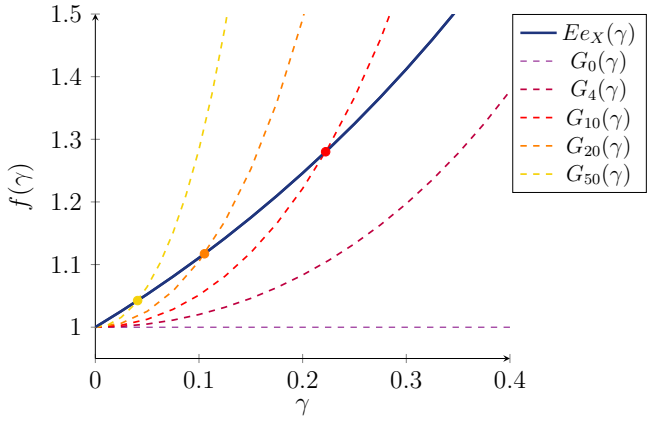


If $E[X] > 0$ then in some neighborhood around $\gamma = 0$ would be greater than any $e^{\frac{c\gamma^2}{2}}$. To proof it, consider function $G_c(\gamma) = e^{\frac{c\gamma^2}{2}}$, then

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

$g(0) = 0 < E[X]$ 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

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

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

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

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

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

3.2 Lower bounds

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

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

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

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

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

which also implies

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

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

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

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

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

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...