

Probabilistic analysis of vantage point trees

Vladyslav Bohun

*Faculty of Computer Science and Cybernetics, Taras Shevchenko National
University of Kyiv, 01601 Kyiv, Ukraine*

vladyslavbogun@gmail.com (V. Bohun)

Received: 27 April 2021, Revised: 8 July 2021, Accepted: 8 July 2021,
Published online: 4 August 2021

Abstract Probabilistic properties of vantage point trees are studied. A vp-tree built from a sequence of independent identically distributed points in $[-1, 1]^d$ with the ℓ_∞ -distance function is considered. The length of the leftmost path in the tree, as well as partitions over the space it produces are analyzed. The results include several convergence theorems regarding these characteristics, as the number of nodes in the tree tends to infinity.

Keywords Fixed-point equation, machine learning, Markov chain, nearest neighbor search, probabilistic analysis, random tree, similarity search, vantage point tree, vp-tree

2010 MSC [60F05](#), [60J05](#), [60J20](#), [68P05](#), [68W40](#)

1 Introduction

A vantage point tree (vp-tree) is a data structure for fast executing of nearest neighbor search queries in a given metric space. This class of trees was first introduced in 1993 in [17], and is widely used since then. It is not the only class of trees used for nearest neighbor search, other famous examples being kd-trees [3, 12], ball trees [14] as well as many other data structures.

Nearest neighbor search is used nowadays in a wide range of tasks. A simple example of day-to-day usage is geographic search of objects close to a specific location. The nearest points search is used in computational geometry, for example, for identifying intersections, which is a routine task in computer graphics or physics simulations. Last but not least example we mention here is a similarity search, such as search for similar images or similar articles.

© 2021 The Author(s). Published by VTeX. Open access article under the [CC BY](#) license.

Another major field, where the nearest neighbor search is an important tool, is machine learning. More precisely, it is used in variety of regression models. We refer the reader to [5] and [7] for a nice introduction to such models. We also refer to [6], which is a good exposition of ideas behind solving classification problem with the nearest neighbors search. These models proved useful in areas of classification such as text categorization, multimedia categorization and search, pattern recognition, information retrieval and many others. Recent studies have started to combine the nearest neighbor search and neural networks for better and more stable results, see, for example, [15].

Although the fast nearest neighbor search is important by itself, as well as an important ingredient in many algorithms, it and the corresponding data structures are not fully understood from the mathematical viewpoint. Most of the results describing the structure of these trees were obtained using empirical methods. For example, in [9] or [16] vp-trees and their heuristic variations were statistically investigated on specific datasets. Another approach is to compare different tree structures, for example, vp-trees and kd-trees, and collate their characteristics on the same datasets, see [11, 13]. In this paper we use a probabilistic approach to analyze a vp-tree and some related characteristics, as the number of nodes in the tree goes to infinity.

The structure of this paper is the following. In Section 2 we describe a vp-tree model and introduce characteristics that we are going to study. Section 3 contains our main results. In Section 4 we specialize some of the results to a number of partial cases and analyze them in more details.

2 Random vantage point tree model

Let $(\mathcal{X}, \text{dist})$ be an arbitrary metric space. A vp-tree is a binary tree in which each node T stores two values: an element $x_T \in \mathcal{X}$ called a vantage point and a threshold $\tau_T > 0$. One of the most common ways to construct a vp-tree is to do this in a recursive manner from a sequence of elements x_1, x_2, \dots . The tree is empty at the beginning, the first element x_1 is always stored in its root and some threshold τ is assigned. In order to insert the n -th element x_n the following recursive procedure is executed. We start from node T which is a root of the tree. If $\text{dist}(x_T, x_n) \leq \tau$ then we proceed recursively by adding a new element to the T -th left son, otherwise we proceed into the right son. When on some step the node is empty, then the value x_n is stored in it and some threshold is assigned.

In this article we study a special class of random vp-trees. More precisely we consider the metric space $([-1, 1]^d, \text{dist})$ ($d \geq 1$) with a max-distance function

$$\text{dist}(x, y) = \max_{1 \leq i \leq d} |x_i - y_i|. \quad (1)$$

We also need to specify the choice of the thresholds. To this end, fix a parameter $\tau \in (0, 1)$. The threshold of each node T is given by the rule $\tau_T = \tau^h$, where h is the distance from T to the root of the tree plus 1. Another possible option is to pick h as the number of “left turns” in the path from the root to T plus 1. Let us emphasize that such a choice of thresholds is quite natural. If one wants to assign thresholds as mean distances then they will also follow the exponential pattern.

We define the *leftmost path* of a vp-tree (or of an arbitrary binary tree) as a path which starts in the root and always goes into the left son of the current node till reaching a leaf. Note that for both aforementioned rules of picking a threshold, the value h is the same for elements of the leftmost path.

The sequence of elements x_1, x_2, \dots is a sequence of independent uniformly distributed points on $[-1, 1]^d$ and this makes the vp-tree, that we construct, random.

Each node in the leftmost path can be enumerated starting from 1. To understand how leftmost path is formed and evolves let us associate with each node T_h ($h \geq 1$) in this path a part of space $I_h \subset \mathcal{X}$, which contains all points x that, if added to the tree would be saved in left subtree of T_h . It is obvious that $I_h \subset I_{h-1}$ and, furthermore, a point x is saved in the left subtree of T_h if $\text{dist}(x, x_{T_h}) \leq \tau^h$. This gives

$$I_h = I_{h-1} \cap B_{\tau^h}(x_{T_h}), \quad h \geq 1, \tag{2}$$

where $B_{\tau^h}(x_{T_h})$ is a metric ball of radius τ^h and center x_{T_h} , that is,

$$B_{\tau^h}(x_{T_h}) = \{x \in \mathcal{X} : \text{dist}(x, x_{T_h}) \leq \tau^h\}. \tag{3}$$

Note that by our choice of the metric, $\text{dist}(x, x_{T_h}) \leq \tau^h$ is a d -dimensional cube. The root T_1 of the tree is always a part of the leftmost path, hence we can naturally put $I_0 := \mathcal{X} = [-1, 1]^d$ which is also a cube. From (2) we conclude that, for all $h \geq 1$, I_h is an intersection of a rectangle and a cube, and hence is a rectangle itself with edges being parallel to coordinate axes. For such figures it is enough (up to parallel translations) to store only information about their sizes in each dimension which we denote by $I_h^{(i)}$, $1 \leq i \leq d$.

Let T_{h-1} be currently the last node in the leftmost path. Then the new node T_h will be added to the left of T_{h-1} if a new point x_n falls inside I_{h-1} . From the properties of the uniform distribution, conditionally on $x_n \in I_{h-1}$, x_n has the uniform distribution on I_{h-1} . Taking into account that I_{h-1} is a rectangle, we also obtain that x_n 's position inside I_{h-1} has the uniform distribution along every dimension $i = 1, \dots, d$. Thus, we can write d recursive equations in each dimension for the size of each side of I_h , which is just a size of an intersection of two segments: a segment of length $I_{h-1}^{(i)}$ and a segment of radius τ^h and a center distributed uniformly inside the first segment. We compute the size $I_h^{(i)}$ as the distance between right and left endpoints of the intersection. Let $u_h^{(i)}$, $h \geq 1$, $1 \leq i \leq d$, be a family of independent uniformly distributed random variables on $[0, 1]$. We have, for $h \geq 1$, $1 \leq i \leq d$,

$$\begin{aligned} I_h^{(i)} &= \min\{I_{h-1}^{(i)}, u_h^{(i)} I_{h-1}^{(i)} + \tau^h\} - \max\{0, u_h^{(i)} I_{h-1}^{(i)} - \tau^h\} \\ &= (u_h^{(i)} I_{h-1}^{(i)} - \max\{0, u_h^{(i)} I_{h-1}^{(i)} - \tau^h\}) \\ &\quad + (\min\{I_{h-1}^{(i)}, u_h^{(i)} I_{h-1}^{(i)} + \tau^h\} - u_h^{(i)} I_{h-1}^{(i)}) \\ &= (u_h^{(i)} I_{h-1}^{(i)} + \min\{0, \tau^h - u_h^{(i)} I_{h-1}^{(i)}\}) \\ &\quad + (\min\{I_{h-1}^{(i)}, u_h^{(i)} I_{h-1}^{(i)} + \tau^h\} - u_h^{(i)} I_{h-1}^{(i)}) \\ &= \min\{u_h^{(i)} I_{h-1}^{(i)}, \tau^h\} + \min\{(1 - u_h^{(i)}) I_{h-1}^{(i)}, \tau^h\}. \end{aligned} \tag{4}$$

Making the substitution $X_h^{(i)} := I_h^{(i)}/\tau^h$ we obtain the basic recursive formula describing evolution of I_h in max-distance case:

$$X_h^{(i)} = \min \left\{ \frac{u_h^{(i)} X_{h-1}^{(i)}}{\tau}, 1 \right\} + \min \left\{ \frac{(1 - u_h^{(i)}) X_{h-1}^{(i)}}{\tau}, 1 \right\}. \tag{5}$$

3 Convergence of the length of the leftmost path

3.1 The analysis of the process (X_h)

The process (5) defines sizes of $X_h = I_h/\tau^h$ in different dimensions. $(X_h)_{h \geq 0}$ can be considered as a Markov chain over rectangles (with omitting their positions). We define by $|X_h|$ (or $|I_h|$) a Lebesgue measure of a corresponding rectangle, which can be simply computed knowing sizes of each edge

$$|X_h| = \prod_{i=1}^d X_h^{(i)}. \tag{6}$$

Proposition 3.1. *The Markov chain (5) satisfies*

$$1 \leq X_h^{(i)} \leq 2, \tag{7}$$

for $h \geq 0$ and $1 \leq i \leq d$.

Proof. Let us fix i and prove the statement on h by using induction. For a random vp-tree $I_0 = \mathcal{X} = B_1(0) = [-1, 1]^d$, hence $X_0^{(i)} = I_0^{(i)} = 2$ and the statement holds. Since $\min\{a, b\} \leq b$ from (5) we have

$$X_h^{(i)} \leq 1 + 1 = 2. \tag{8}$$

To show that $X_h^{(i)} \geq 1$ we first note that if any of the minimums in (5) takes value 1, then the statement holds. On the other hand, if both minimums are less than 1, then

$$X_h^{(i)} = \frac{u_h^{(i)} X_{h-1}^{(i)}}{\tau} + \frac{(1 - u_h^{(i)}) X_{h-1}^{(i)}}{\tau} = \frac{X_{h-1}^{(i)}}{\tau} \geq X_{h-1}^{(i)} \geq 1 \tag{9}$$

by the inductive assumption, and since $\tau \in (0, 1)$. □

Let us denote by S a set of all possible values the chain (X_h) can visit provided that $X_0 = B_1(0)$. Note that S is uncountable.

We shall use the notation of stochastic ordering which we recall for convenience.

Definition 3.2. For two random variables X and Y we say that Y stochastically dominates X and write $X \stackrel{st}{\leq} Y$ if

$$\mathbb{P}\{X \leq t\} \geq \mathbb{P}\{Y \leq t\} \quad \text{for all } t \in \mathbb{R}. \tag{10}$$

The key ingredient for the subsequent analysis is the following theorem which says that the chain $(X_h)_{h \geq 0}$ visits the state $B_1(0)$ relatively often.

Theorem 3.3. Fix $\alpha \in S$ and put

$$R_\alpha := \inf\{h \geq 1 : X_h = B_1(0) \text{ given that } X_0 = \alpha\}.$$

Then $\mathbb{E}R_\alpha < \infty$.

Remark 3.4. We do believe that the counterpart of this theorem holds in arbitrary metric on $[-1, 1]^d$, however we have not been able to prove it in such generality.

Proof of Theorem 3.3. We rewrite (2) as follows:

$$X_{h+1} = \frac{X_h}{\tau} \cap B_1(u_{h+1}^*), \tag{11}$$

where u_{h+1}^* is uniformly distributed inside X_h/τ .

Since the positions of rectangles are not important, the recursive equation (11) can be also rewritten as

$$X_{h+1} = \frac{X_h - u_{h+1}}{\tau} \cap B_1(0), \tag{12}$$

where u_{h+1} is uniformly distributed in X_h .

From Proposition 3.1 it follows that for any state X_h there is a d -dimensional square with sides 1 inside X_h , or in other words there is always a point \hat{x}_h such that $B_\delta(\hat{x}_h) \subset X_h$ with $\delta = 1/2$.

Let us assume that point u_{h+1} is chosen inside $B_\delta(\hat{x}_h)$ and denote

$$\Delta = \text{dist}(u_{h+1}, \hat{x}_h) \leq \delta. \tag{13}$$

From the triangle inequality we obtain

$$\max_{y \in \partial B_\delta(\hat{x}_h)} \text{dist}(u_{h+1}, y) \leq \Delta + \delta, \tag{14}$$

$$\min_{y \in \partial B_\delta(\hat{x}_h)} \text{dist}(u_{h+1}, y) \geq \delta - \Delta, \tag{15}$$

where ∂B defines the boundary of the set B .

Then, in view of (12), we have

$$\max_{y \in \partial((B_\delta(\hat{x}_h) - u_{h+1})/\tau)} \text{dist}(0, y) \leq \frac{\Delta + \delta}{\tau}, \tag{16}$$

and

$$\min_{y \in \partial((B_\delta(\hat{x}_h) - u_{h+1})/\tau)} \text{dist}(0, y) \geq \frac{\delta - \Delta}{\tau}. \tag{17}$$

From (16) we have our first property that if $\frac{\Delta + \delta}{\tau} \leq 1$ (or equally $\Delta \leq \tau - \delta$), then $B_{\delta/\tau}((\hat{x}_h - u_{h+1})/\tau) \subset X_{h+1}$. From (17) we have our second property that if $\frac{\delta - \Delta}{\tau} \geq 1$ (or equally $\Delta \leq \delta - \tau$), then $X_{h+1} = B_1(0)$.

Let us now fix a constant $k = \min\{k \in \mathbb{N} : \frac{1+\tau}{2} \tau^{k-1} \leq \frac{1}{2}\}$. Since $B_{1/2}(\hat{x}_h) \subset X_h$, we have $B_{\frac{1+\tau}{2}\tau^{k-1}}(\hat{x}_h) \subset X_h$. Since $k \geq 2$, it holds $\frac{1+\tau}{2} \tau^{k-1} < \tau$. Therefore, if

$u_{h+1} \in B_{\varepsilon_1}(\hat{x}_h)$ with $\varepsilon_1 = \tau - \frac{1+\tau}{2}\tau^{k-1}$, which happens with a positive probability, we will obtain $B_{\frac{1+\tau}{2}\tau^{k-2}}(\hat{x}_{h+1}) \subset X_{h+1}$ from the first property. Repeating this $k - 1$ times in total, where $\varepsilon_i = \tau - \frac{1+\tau}{2}\tau^{k-i} > 0$, for $1 \leq i \leq k - 1$, we obtain that $B_{\frac{1+\tau}{2}}(\hat{x}_{h+k-1}) \subset X_{h+k-1}$ and $\frac{1+\tau}{2} > \tau$. Thus, if $u_{h+k} \in B_{\varepsilon_k}(\hat{x}_{h+k-1})$, which happens with a positive probability, where $\varepsilon_k = \frac{1+\tau}{2} - \tau$, we get that $X_{h+k} = B_1(0)$ from the second property. Summarizing, we see that from an arbitrary state $\alpha \in S$ the chain jumps in k steps to the state $B_1(0)$ with a positive probability which is independent of the starting state α . Note that k also only depends on τ .

We split the evolution of the chain into blocks of length k . The state at the end of each block can be $B_1(0)$ with a positive probability, say p , obtained from the described scheme. We denote by g a number of the first block which ends up in the state $B_1(0)$. Then since actual probability to end up in the state $B_1(0)$ at the end of the block is not less than p we have, for any $t \geq 0$,

$$\mathbb{P}\{g \leq t\} \geq \mathbb{P}\{G \leq t\}, \tag{18}$$

where G has a geometric distribution with parameter p (number of trials). Or in terms of stochastic order

$$g \stackrel{st}{\leq} G. \tag{19}$$

Similarly, since the moment R_α of the first visit to state $B_1(0)$ (after starting state) can occur in the middle of each block, we can write

$$R_\alpha \stackrel{st}{\leq} kg. \tag{20}$$

From the properties of stochastic order, inequality also holds for expectations, hence combining last two comparisons we have

$$\mathbb{E}R_\alpha \leq k\mathbb{E}G = \frac{k}{p} < \infty. \tag{21}$$

□

Remark 3.5. We note that in the proof of Theorem 3.3 we did not significantly use that u_h has the uniform distribution, and it works for any absolutely continuous distribution, since probability for u_h to fall inside some $B_\varepsilon(\hat{x}_h)$ is positive and larger than some fixed positive constant.

Remark 3.6. The process (11) in particular case $\tau = 1$ is called the diminishing process of Balint Toth. This process was studied in various partial cases including one-dimensional [1] and multi-dimensional case [10] with $B_1(0) = [-1, 1]^d$, as well as in other metrics [2, 10]. However evolution of this process with $\tau < 1$ differs a lot and requires an independent study, whilst methods used for $\tau = 1$ cannot be applied to treat the case $\tau < 1$.

From Theorem 3.3 it follows that the chain $(X_h)_{h \geq 0}$ evolves as follows. There is a special state $B_1(0)$, which the chain visits infinitely often and time to return to this state is a.s. finite and, moreover, has a finite mean. Thus, the whole evolution of X_h can be split into independent cycles between visiting the state $B_1(0)$.

In order to proceed we need to recall a notion of the Harris chains. For information about Harris chains and examples we refer the reader to the book [8, Chapter 6]. We will only present here the required definitions.

Definition 3.7. A Markov chain $(X_h)_{h \geq 0}$ on a state space S is a Harris chain if there are two sets $A, B \subset S$, a positive function $q(x, y) \geq \varepsilon > 0$ for $x \in A, y \in B$, and a probability measure ρ concentrated on B such that the following two conditions hold:

1. $\mathbb{P}\{\inf\{h \geq 0 : X_h \in A\} < \infty\} > 0$ for all possible starting states $X_0 \in S$;
2. if $x \in A$ and $C \subset B$, then $\mathbb{P}\{X_{h+1} \in C | X_h = x\} \geq \int_C q(x, y)\rho(dy)$.

Definition 3.8. A Harris chain is called recurrent if there is a state α such that $\mathbb{P}\{\inf\{h \geq 1 : X_h = \alpha\} < \infty\} = 1$ when $X_0 = \alpha$.

Definition 3.9. A recurrent Harris chain is aperiodic if the greatest common divisor of set $\{h \geq 1 : \mathbb{P}\{X_h = \alpha | X_0 = \alpha\} > 0\}$ is 1.

Proposition 3.10. $(X_h)_{h \geq 0}$ is an aperiodic recurrent Harris chain.

Proof. Let us first show that $(X_h)_{h \geq 0}$ is a Harris chain. Put $A = B = \{B_1(0)\}$. From Theorem 3.3 it follows that $\mathbb{P}\{R_{X_0} < \infty\} = 1$. Since $\inf\{h \geq 0 : X_h \in A\} < \infty \leq R_{X_0}$ for all possible starting states, the first condition holds. For the second part of the definition we put $q(x, y) := p := \mathbb{P}\{X_{h+1} = B_1(0) | X_h = B_1(0)\}$. From the proof of Theorem 3.3, more precisely from the property (17), it follows that $p > 0$.

To show recurrence of the chain we put $\alpha = B_1(0)$. Then from Theorem 3.3 we have that $R_\alpha < \infty$ almost surely. For the aperiodicity it is enough to note that $\mathbb{P}\{X_1 = \alpha | X_0 = \alpha\} = p > 0$. □

Theorem 3.11. The Markov chain $(X_h)_{h \geq 0}$ has a stationary distribution I_∞ and

$$\lim_{n \rightarrow \infty} \|\mathbb{P}\{X_h = \cdot | X_0 = B_1(0)\} - I_\infty(\cdot)\|_{TV} = 0, \tag{22}$$

where $\|\cdot\|_{TV}$ denotes the total variance distance.

Proof. From Proposition 3.10 we know that (X_h) is an aperiodic recurrent Harris chain. By Theorem 6.8.5 in [8, p. 323] this chain has a stationary measure. By Theorem 6.8.8 in [8, p. 324] an aperiodic recurrent Harris chain converges to a stationary distribution in the total variation distance as (22) if $R_{B_1(0)}$ is almost surely finite. The latter is secured by Theorem 3.3. □

3.2 Limit theorems for the length of the leftmost path

Using the results on the behavior of the chain (X_h) we can analyze the length of the leftmost path of a vp-tree constructed from n random points. Let us introduce the following random walk

$$S_0 := 0, \quad S_k := \sum_{h=0}^{k-1} G_h \quad k \geq 1, \tag{23}$$

where, given the path of $(I_h)_{h \geq 0}$, $(G_h)_{h \geq 0}$ are conditionally independent random variables such that the distribution of G_h is geometric on \mathbb{N} with parameter $\frac{|I_h|}{2^d}$. This

sum shows how many points are needed to have the leftmost path of size at least k , since the new node is added to it when a point x_n falls into area of size $|I_h|$ which is, given $(I_h)_{h \geq 0}$, a Bernoulli trial.

The size (number of nodes) of the leftmost path in terms of S_k can be expressed as

$$L_n = \max\{k \geq 0 : S_k \leq n\}. \tag{24}$$

The length of the leftmost path can be simply computed from its size as $L_n - 1$.

Theorem 3.12. *For a random vp-tree the size of the leftmost path L_n satisfies the following weak law of large numbers, as $n \rightarrow \infty$,*

$$\frac{L_n}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{-\log \tau}. \tag{25}$$

To prove this theorem we need two lemmas. We note that if $X \sim \text{Geom}(a)$ and $Y \sim \text{Geom}(b)$, then

$$X \stackrel{st}{\leq} Y \Leftrightarrow a \geq b. \tag{26}$$

Lemma 3.13. *For random variables $G_k \sim \text{Geom}(\frac{|I_k|}{2^d})$ ($k \geq 0$)*

$$\frac{1}{k} \log G_k \xrightarrow{\mathbb{P}} -\log \tau, \tag{27}$$

as $k \rightarrow \infty$.

Proof. We compute the distribution function of $\frac{1}{k} \log G_k$. Note that

$$\mathbb{P} \left\{ \frac{1}{k} \log G_k \leq x \right\} = \mathbb{E} \left\{ \mathbb{P} \left\{ \frac{1}{k} \log G_k \leq x \mid I_k \right\} \right\}, \tag{28}$$

and further

$$\mathbb{P}\{k^{-1} \log G_k \leq x \mid I_k\} = \mathbb{P}\{G_k \leq \exp(kx) \mid I_k\} = 1 - (1 - |I_k|/2^d)^{\lfloor \exp(kx) \rfloor}. \tag{29}$$

From Proposition 3.1 we have that

$$\tau^k \leq |I_k| \leq 2^d \tau^k. \tag{30}$$

Thus we can bound the probability in (29) by inserting lower and upper bounds of $|I_k|$. Thus,

$$1 - (1 - 2^{-d} \tau^k)^{\lfloor \exp(kx) \rfloor} \leq \mathbb{P}\{k^{-1} \log G_k \leq x \mid I_k\} \leq 1 - (1 - \tau^k)^{\lfloor \exp(kx) \rfloor}.$$

Both upper and lower bounds converge, as $k \rightarrow \infty$, to

$$\begin{cases} 0, & \text{if } \log \tau + x < 0 \\ 1, & \text{if } \log \tau + x > 0 \end{cases}. \tag{31}$$

Hence by the standard sandwich argument $\mathbb{P}\{k^{-1} \log G_k \leq x \mid I_k\}$ converges to the same values almost surely. By the dominated convergence theorem, the same holds for the unconditional probability $\mathbb{P}\{\frac{1}{k} \log G_k \leq x\}$. This finishes the proof. \square

Lemma 3.14. For a random walk S_k described in (23)

$$\frac{1}{k} \log S_k \xrightarrow{\mathbb{P}} -\log \tau \tag{32}$$

as $k \rightarrow \infty$.

Proof. We write the distribution function of S_k as

$$\mathbb{P}\{S_k \leq x\} = \mathbb{E}\{\mathbb{P}\{S_k \leq x | (I_h)_{0 \leq h < k}\}\} \tag{33}$$

and study the conditional probability under the expectation.

Since $S_k \geq G_{k-1}$ we have an upper bound

$$\mathbb{P}\{S_k \leq x | (I_h)_{0 \leq h < k}\} \leq \mathbb{P}\{G_{k-1} \leq x | (I_h)_{0 \leq h < k}\}, \tag{34}$$

and the inequality holds for expectations, that is, the unconditional probability.

Let us recall that, given $(I_h)_{0 \leq h \leq k}$, the random variables $(G_h)_{0 \leq h \leq k}$ are independent. We also have that $G_h \stackrel{st}{\leq} G_k$ ($0 \leq h < k$) since the sequence $(|I_h|)_{h \geq 0}$ is nonincreasing. Thus, we have the bound

$$\sum_{h=0}^{k-1} G_h \stackrel{st}{\leq} \sum_{h=0}^{k-1} G_k^{(h)} \stackrel{st}{\leq} k \max_{0 \leq h < k} G_k^{(h)} = k \hat{G}_k \tag{35}$$

where $(G_k^{(h)})_{0 \leq h < k}$ is a set of independent copies of geometrically distributed random variables with the parameter $\tau^k / 2^d$, providing $G_k \stackrel{st}{\leq} G_k^{(1)}$, and we put $\hat{G}_k := \max_{0 \leq h < k} G_k^{(h)}$. Note that the inequality stays for expectations, giving that it holds for unconditional values of S_k .

We compute the distribution function of \hat{G}_k :

$$\mathbb{P}\left\{\frac{1}{k} \log \hat{G}_k \leq x\right\} = \mathbb{P}\{\hat{G}_k \leq \exp(kx)\} = \left(1 - (1 - \tau^k / 2^d)^{\lfloor \exp(kx) \rfloor}\right)^k, \tag{36}$$

and therefore, as $k \rightarrow \infty$,

$$\mathbb{P}\left\{\frac{1}{k} \log \hat{G}_k \leq x\right\} \rightarrow \begin{cases} 0, & \text{if } \log \tau + x < 0 \\ 1, & \text{if } \log \tau + x \geq 0 \end{cases}. \tag{37}$$

Now we combine (34) and (35) to obtain

$$\log G_{k-1} \stackrel{st}{\leq} \log S_k \stackrel{st}{\leq} \log k + \log \hat{G}_k. \tag{38}$$

Dividing everything by k we obtain

$$\frac{k-1}{k} \frac{\log G_{k-1}}{k-1} \stackrel{st}{\leq} \frac{\log S_k}{k} \stackrel{st}{\leq} \frac{\log k}{k} + \frac{\log \hat{G}_k}{k}. \tag{39}$$

From Lemma 3.13 and (37) sending $k \rightarrow \infty$ yields

$$\frac{1}{k} \log S_k \xrightarrow{d} -\log \tau. \tag{40}$$

Convergence in probability is implied from the fact that the limit is a constant. \square

Proof of Theorem 3.12. We start by computing the distribution function of $\frac{L_n}{\log n}$ in any point $x > 0$ and $x \neq -1/\log \tau$:

$$\begin{aligned} \mathbb{P} \left\{ \frac{L_n}{\log n} > x \right\} &= \mathbb{P}\{L_n > x \log n\} = \mathbb{P}\{S_{\lfloor x \log n \rfloor} \leq n\} \\ &= \mathbb{P}\{\log S_{\lfloor x \log n \rfloor} \leq \log n\} = \mathbb{P} \left\{ \frac{\log S_{\lfloor x \log n \rfloor}}{\log n} \leq 1 \right\} \\ &= \mathbb{P} \left\{ \frac{\log S_{\lfloor x \log n \rfloor}}{x \log n} \leq \frac{1}{x} \right\}. \end{aligned} \tag{41}$$

By Lemma 3.14 we have

$$\mathbb{P} \left\{ \frac{\log S_{\lfloor x \log n \rfloor}}{x \log n} \leq \frac{1}{x} \right\} \rightarrow \begin{cases} 0, & 1/x < -\log \tau \\ 1, & 1/x > -\log \tau \end{cases}, \tag{42}$$

as $k \rightarrow \infty$, and therefore

$$\mathbb{P} \left\{ \frac{L_n}{\log n} \leq x \right\} = 1 - \mathbb{P} \left\{ \frac{L_n}{\log n} > x \right\} \rightarrow \begin{cases} 1, & x > -1/\log \tau \\ 0, & x < -1/\log \tau \end{cases}. \tag{43}$$

Since the convergence in distribution to a constant implies the convergence in probability to the same constant, the theorem is proved. \square

We note that for proving Theorem 3.12 we only used that $|I_h|/\tau^h$ is bounded both from zero and infinity by some absolute constants. The following results demonstrate convergence of L_n over subsequences. We first state and prove the result for the random walk (S_k) . Recall that I_∞ is the stationary distribution of the chain $(I_h/\tau^h)_{h \geq 0}$.

Lemma 3.15. *For a random walk S_k defined by (23), as $k \rightarrow \infty$,*

$$\tau^k S_k \xrightarrow{d} G_\infty, \tag{44}$$

where G_∞ is a random series $\sum_{j=1}^\infty \xi_j$. Here $(\xi_j)_{j \geq 0}$ are conditionally independent, given I_∞ , random variables such that $\xi_j \sim \text{Exp}(\frac{|I_\infty|}{2^d \tau^j})$.

Proof. Let us write the characteristic function of $\tau^k S_k$:

$$\begin{aligned} \mathbb{E} e^{it\tau^k S_k} &= \mathbb{E} \left[\mathbb{E} \left(e^{it\tau^k S_k} | (I_h)_{0 \leq h < k} \right) \right] = \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{|I_j|/2^d}{e^{-it\tau^k} - (1 - |I_j|/2^d)} \right] \\ &= \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{|I_j|/2^d}{1 - it\tau^k + O(\tau^{2k}) - (1 - |I_j|/2^d)} \right] \\ &= \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{|I_j|/2^d}{|I_j|/2^d - it\tau^k + O(\tau^{2k})} \right] \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{E} \left[\prod_{j=0}^{k-1} \frac{1}{1 - 2^d i t \tau^{k-j} \frac{\tau^j}{|I_j|} + O\left(\frac{\tau^{2k}}{|I_j|}\right)} \right] \\
 &= \mathbb{E} \left[\prod_{j=1}^k \frac{1}{1 - 2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O\left(\frac{\tau^k}{|I_{k-j}|}\right) O(\tau^k)} \right] \\
 &= \mathbb{E} \left[\prod_{j=1}^M \frac{1}{1 - 2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k)} \cdot \prod_{j=M+1}^k \frac{1}{1 - 2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k)} \right] \\
 &= \mathbb{E} [A_{k,M}(t) \cdot B_{k,M}(t)], \tag{45}
 \end{aligned}$$

where $1 \leq M < k$ is a fixed parameter and $O\left(\frac{\tau^k}{|I_{k-j}|}\right) = O(1)$ by Proposition 3.1.

From Theorem 3.11 the convergence in total variance implies the convergence in distribution, the convergence of modules is also implied considering specifics of given rectangles. Thus, for every fixed M as $k \rightarrow \infty$ we have

$$A_{k,M}(t) \xrightarrow{d} \prod_{j=1}^M \frac{1}{1 - 2^d i t \tau^j / |I_{\infty}|}. \tag{46}$$

Let us show now that, for any $\varepsilon > 0$,

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} \mathbb{P}\{|A_{k,M}(t) B_{k,M}(t) - A_{k,M}(t)| \geq \varepsilon\} = 0. \tag{47}$$

By Markov’s inequality

$$\mathbb{P}\{|A_{k,M}(t) B_{k,M}(t) - A_{k,M}(t)| \geq \varepsilon\} \leq \frac{\mathbb{E}|A_{k,M}(t)| |B_{k,M}(t) - 1|}{\varepsilon}. \tag{48}$$

At first we want to show that $B_{k,M}(t)$ converges to 1 by separately showing that module and argument of the complex value converge. We have

$$\begin{aligned}
 |B_{k,M}(t)|^2 &= \prod_{j=M+1}^k \frac{1}{(1 + O(\tau^k))^2 + (2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k))^2}, \tag{49} \\
 \log |B_{k,M}(t)|^2 &= - \sum_{j=M+1}^k \log \left[(1 + O(\tau^k))^2 + \left(2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k) \right)^2 \right] \\
 &= - \sum_{j=M+1}^k \log \left[1 + O(\tau^k) + 2^{2d} i^2 \tau^{2j} \left(\frac{\tau^{k-j}}{|I_{k-j}|} \right)^2 + O(\tau^k) \right] \\
 &= - \sum_{j=M+1}^k \left[2^{2d} i^2 \tau^{2j} \left(\frac{\tau^{k-j}}{|I_{k-j}|} \right)^2 + O(\tau^k) + O(\tau^{4j}) + O(\tau^{2j+k}) + O(\tau^{2k}) \right] \\
 &= - \sum_{j=M+1}^k [O(\tau^{2j}) + O(\tau^k)].
 \end{aligned} \tag{50}$$

The expansion for logarithm holds starting with some large enough M . We also used Proposition 3.1 which shows that $\tau^{k-j}/|I_{k-j}|$ is bounded by nonrandom constants, which also do not dependent on $k - j$. We note that for a fixed t , the constants in the remainders $O(\tau^{2j})$ and $O(\tau^k)$ are uniform, that is, do not depend on j, k . Thus, we can write

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} \sum_{j=M+1}^k \left[O(\tau^{2j}) + O(\tau^k) \right] = 0 \quad \text{a.s.} \tag{51}$$

For the argument we write

$$\begin{aligned} \arg B_{k,M}(t) &= - \sum_{j=M+1}^k \text{Arg} \left(1 - 2^d i t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|} + O(\tau^k) \right) \\ &= \sum_{j=M+1}^k \left[\frac{2^d t \tau^j \frac{\tau^{k-j}}{|I_{k-j}|}}{\sqrt{1 + 2^{2d} t^2 \tau^{2j} \left(\frac{\tau^{k-j}}{|I_{k-j}|} \right)^2}} + O(\tau^{3j}) + O(\tau^k) \right] \\ &= \sum_{j=M+1}^k \left[O(\tau^j) + O(\tau^k) \right]. \end{aligned} \tag{52}$$

The expansion for the principal argument holds starting from some large enough M . Arguing as before we can write

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} \arg B_{k,M}(t) = 0 \quad \text{a.s.} \tag{53}$$

We also showed in process that starting from some large enough k and M the quantity $|B_{k,M}(t)|$ is bounded by a nonrandom constant $\bar{B} > 0$ on every stochastic path. Since $|A_{k,M}(t)B_{k,M}(t)| = 1$, then $|A_{k,M}(t)|$ is also bounded by some constant $\bar{A} > 0$. Thus, we can write

$$\frac{\mathbb{E}|A_{k,M}(t)||B_{k,M}(t) - 1|}{\varepsilon} \leq \frac{\bar{A}\mathbb{E}|B_{k,M}(t) - 1|}{\varepsilon}. \tag{54}$$

We already showed that

$$\lim_{M \rightarrow \infty} \limsup_{k \rightarrow \infty} |B_{k,M}(t) - 1| = 0 \quad \text{a.s.}$$

Since $|B_{k,M}(t) - 1| \leq 1 + \bar{B}$, then by the dominated convergence theorem

$$\lim_{M \rightarrow \infty} \lim_{k \rightarrow \infty} \mathbb{E}|B_{k,M}(t) - 1| = 0. \tag{55}$$

Combining pieces together we obtain (47).

By Theorem 3.2 in [4, p. 28] we have that as $k \rightarrow \infty$

$$e^{i t \tau^k S_k} \xrightarrow{d} \lim_{M \rightarrow \infty} \prod_{j=1}^M \frac{1}{1 - 2^d i t \tau^j / |I_{\infty}|} = \prod_{j=1}^{\infty} \frac{1}{1 - 2^d i t \tau^j / |I_{\infty}|}. \tag{56}$$

Elements of the sequence $(e^{it\tau^k S_k})_{k \geq 0}$ are uniformly integrable as characteristic functions, so expectations also converge as $k \rightarrow \infty$,

$$\mathbb{E}e^{it\tau^k S_k} \rightarrow \mathbb{E} \prod_{j=1}^{\infty} \frac{1}{1 - 2^d i t \tau^j / |I_{\infty}|}, \tag{57}$$

and the limiting value here is a characteristic function of G_{∞} which finishes the proof. \square

Theorem 3.16. *For any fixed $T > 0$,*

$$L_{\lfloor T\tau^{-n} \rfloor} - n \xrightarrow{d} L_{\infty}^{(T)} \tag{58}$$

as $n \rightarrow \infty$ and the distribution function of $L_{\infty}^{(T)}$ is given by

$$\mathbb{P}\{L_{\infty}^{(T)} \leq m\} = 1 - \mathbb{P}\{G_{\infty} \leq T\tau^m\} \tag{59}$$

for any $m \in \mathbb{Z}$ with G_{∞} defined in Lemma 3.15.

Proof. Since $L_{\lfloor T\tau^{-n} \rfloor} - n$ takes values only on \mathbb{Z} , we study the distribution function only over integer values. For fixed $T > 0$ and $m \in \mathbb{Z}$ we have

$$\begin{aligned} \mathbb{P}\{L_{\lfloor T\tau^{-n} \rfloor} - n > m\} &= \mathbb{P}\{L_{\lfloor T\tau^{-n} \rfloor} > n + m\} = \mathbb{P}\{S_{n+m} \leq \lfloor T\tau^{-n} \rfloor\} \\ &= \mathbb{P}\{\tau^{n+m} S_{n+m} \leq \tau^{n+m} \lfloor T\tau^{-n} \rfloor\}. \end{aligned} \tag{60}$$

From Lemma 3.15 and the fact that $\tau^n \lfloor T\tau^{-n} \rfloor \rightarrow T$ as $n \rightarrow \infty$ we obtain

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\tau^{n+m} S_{n+m} \leq \tau^{n+m} \lfloor T\tau^{-n} \rfloor\} = \mathbb{P}\{G_{\infty} \leq T\tau^m\}, \tag{61}$$

which finishes the proof. \square

4 One-dimensional convergence

In this section we analyze the random recursive equation (5). Sending $h \rightarrow \infty$ in (5) we formally obtain the stochastic fixed-point equation for the limit X_{∞} of $X_h^{(i)}$:

$$X_{\infty} \stackrel{d}{=} \min\{\tau^{-1}uX_{\infty}, 1\} + \min\{\tau^{-1}(1-u)X_{\infty}, 1\}, \tag{62}$$

where u is a uniformly distributed random variable on $[0, 1]$. Let us show, using the contraction method, that $X_h^{(i)}$ converges to X_{∞} , as $h \rightarrow \infty$.

Theorem 4.1. *The function*

$$f(X, u) = \text{Law} \left(\min \left\{ \frac{uX}{\tau}, 1 \right\} + \min \left\{ \frac{(1-u)X}{\tau}, 1 \right\} \right) \tag{63}$$

is a strict contraction in the space of probability measures on $[1, 2]$ endowed with the ℓ_1 -minimal metric, where u is a uniformly distributed random variable on $[0, 1]$ independent of a random variable X taking values in $[1, 2]$.

Definition 4.2. The ℓ_1 -minimal metric between two random variables X and Y is given by

$$\ell_1(X, Y) = \inf \mathbb{E}|\hat{X} - \hat{Y}|, \tag{64}$$

where the infimum is taken over all pairs of random variables (\hat{X}, \hat{Y}) such that $\hat{X} \sim X$ and $\hat{Y} \sim Y$.

Proof of Theorem 4.1. Similarly to the proof of Proposition 3.1 it can be checked that if X takes values in $[1, 2]$, the same holds also for $f(X, u)$.

We prove the contraction property in two steps.

Step 1. We compute the ℓ_1 -minimal distance between $f(x, u_1)$ and $f(y, u_2)$ for fixed nonrandom numbers $x, y \in [1, 2]$. We have

$$\ell_1(f(x, u), f(y, u)) = \inf \mathbb{E}|f(x, u_1) - f(x, u_2)|. \tag{65}$$

Since the infimum is taken over all possible dependencies between u_1 and u_2 , we may pick $u_1 = u_2 = u$. Then

$$\ell_1(f(x, u_1), f(y, u_2)) \leq \int_{[0,1]} |f(x, u) - f(y, u)|du. \tag{66}$$

Let us assume for the time being that there is $\alpha \in (0, 1)$ such that

$$\int_{[0,1]} |f(x, u) - f(y, u)|du \leq \alpha|x - y|. \tag{67}$$

Step 2. We compute the ℓ_1 -minimal distance between $f(X, u_1)$ and $f(Y, u_2)$. Combining the above estimates, we obtain

$$\ell_1(f(X, u_1), f(Y, u_2)) \leq \alpha \mathbb{E}|\hat{X} - \hat{Y}|, \tag{68}$$

for arbitrary pair (\hat{X}, \hat{Y}) such that $\hat{X} \sim X$ and $\hat{Y} \sim Y$. Passing to the infimum over all such pairs, we obtain

$$\ell_1(f(X, u_1), f(Y, u_2)) \leq \alpha \ell_1(X, Y) \tag{69}$$

for some constant $\alpha \in (0, 1)$, which is the definition of contraction.

It remains to prove (67). At first we divide f into sum of two functions.

$$f(x, u) = f_1(x, u) + f_2(x, u), \tag{70}$$

where

$$f_1(x, u) = \min \left\{ \frac{uX}{\tau}, 1 \right\} = \begin{cases} \frac{uX}{\tau}, & u < \frac{\tau}{x} \\ 1, & u \geq \frac{\tau}{x} \end{cases} \tag{71}$$

and $f_2(x, u) = \min\{\frac{(1-u)X}{\tau}, 1\}$.

We compute the following integral under the condition $x \leq y$:

$$\int_0^1 |f_1(x, u) - f_1(y, u)|du = \int_0^{\tau/y} \left(\frac{uy}{\tau} - \frac{ux}{\tau} \right) du$$

$$\begin{aligned}
 & + \int_{\tau/y}^{\tau/x} \left(1 - \frac{ux}{\tau}\right) du + \int_{\tau/x}^1 (1 - 1) du \\
 = & \int_0^{\tau/y} \frac{u}{\tau}(y - x) du + \int_{\tau/y}^{\tau/x} du - \int_{\tau/y}^{\tau/x} \frac{x}{\tau} du \\
 = & \frac{y - x}{\tau} \frac{u^2}{2} \Big|_0^{\tau/y} + \left(\frac{\tau}{x} - \frac{\tau}{y}\right) - \frac{x}{\tau} \frac{u^2}{2} \Big|_{\tau/y}^{\tau/x} \\
 = & \frac{y - x}{2\tau} \frac{\tau^2}{y^2} + \left(\frac{\tau}{x} - \frac{\tau}{y}\right) - \frac{x}{2\tau} \left(\frac{\tau^2}{x^2} - \frac{\tau^2}{y^2}\right) \\
 = & \frac{\tau(y - x)}{2y^2} + \frac{\tau}{x} - \frac{\tau}{y} - \frac{\tau}{2x} + \frac{\tau x}{2y^2} \\
 = & \frac{\tau x(y - x) + 2\tau y^2 - 2\tau xy - \tau y^2 + \tau x^2}{2xy^2} \\
 = & \frac{-\tau xy + \tau y^2}{2xy} = \frac{\tau}{2} \frac{y - x}{xy}.
 \end{aligned} \tag{72}$$

Similarly if $y \leq x$ we obtain that

$$\int_0^1 |f_1(x, u) - f_1(y, u)| du \leq \frac{\tau}{2} \frac{x - y}{xy}, \tag{73}$$

which results in

$$\int_0^1 |f_1(x, u) - f_1(y, u)| du \leq \frac{\tau}{2} \frac{|x - y|}{xy}. \tag{74}$$

Now we estimate the same integral for f_2 , where

$$f_2(x, u) = \begin{cases} 1, & 1 - u \geq \frac{\tau}{x} \\ \frac{(1-u)x}{\tau}, & 1 - u < \frac{\tau}{x} \end{cases}. \tag{75}$$

In the integral we make a substitution $v = 1 - u$:

$$\int_0^1 |f_2(x, u) - f_2(y, u)| du = \int_0^1 |f_2(x, 1 - v) - f_2(y, 1 - v)| dv, \tag{76}$$

where we instantly obtain the same estimation as (74), since $f_2(x, 1 - v) = f_1(x, v)$.

Finally,

$$\begin{aligned}
 \int_{[0,1]} |f(x, u) - f(y, u)| du & \leq \int_0^1 |f_1(x, u) - f_1(y, u)| du \\
 & + \int_0^1 |f_2(x, u) - f_2(y, u)| du \\
 & \leq \frac{\tau}{2} \frac{|x - y|}{xy} + \frac{\tau}{2} \frac{|x - y|}{xy} = \tau \frac{|x - y|}{xy} \leq \tau |x - y|,
 \end{aligned} \tag{77}$$

which proves (67) with $\alpha = \tau$. □

Corollary 4.3. *The stochastic fixed-point equation (62) has the unique solution X_∞ in the space of probability measures on $[1, 2]$ and $X_h^{(i)} \xrightarrow{d} X_\infty$, as $h \rightarrow \infty$, for every fixed $i = 1, \dots, d$.*

Remark 4.4. We obtained the same result as in the previous section without resort to geometric interpretation but with explicit usage of the distribution of sequence of points x_1, x_2, \dots . It also allowed us to deduce the stochastic fixed-point equation for X_∞ . Recall that I_∞ is a rectangle with edges being parallel to coordinate axes, its sizes in all dimensions can be given by $(X_\infty^{(1)}, \dots, X_\infty^{(d)})$, where $(X_\infty^{(i)})_{1 \leq i \leq d}$ are independent copies of X_∞ .

It turns out that the distribution of X_∞ can be found explicitly using the stochastic fixed-point equation (62) by directly computing its distribution function $F(t) = \mathbb{P}\{X_\infty < t\}$. The solution depends heavily on the parameter τ .

We start with a rather simple and special example $\tau \leq 1/2$.

For $t \in [1, 2]$ we write

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= \mathbb{P}\left\{\min\left\{\frac{uX_\infty}{\tau}, 1\right\} + \min\left\{\frac{(1-u)X_\infty}{\tau}, 1\right\} < t\right\} \\ &= \mathbb{P}\left\{1 + 1 < t, \frac{uX_\infty}{\tau} \geq 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &\quad + \mathbb{P}\left\{\frac{uX_\infty}{\tau} + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} < 1\right\} \\ &\quad + \mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &\quad + \mathbb{P}\left\{1 + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} \geq 1, \frac{(1-u)X_\infty}{\tau} < 1\right\}. \end{aligned} \tag{78}$$

The first term is obviously 0, since we assume $t \leq 2$. The second term is also 0 since events $\{\frac{uX_\infty}{\tau} < 1\}$ and $\{\frac{(1-u)X_\infty}{\tau} < 1\}$ cannot occur at the same time for $\tau \leq 1/2$. And the last two terms are symmetrical.

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= 2\mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= 2\mathbb{P}\left\{\frac{uX_\infty}{\tau} + 1 < t, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= 2\mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}, u \leq \frac{X_\infty - \tau}{X_\infty}\right\} = 2\mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}\right\}. \end{aligned} \tag{79}$$

Here in the second equality we used that the event $\{\frac{uX_\infty}{\tau} + 1 < t\}$ implies the event $\{\frac{uX_\infty}{\tau} < 1\}$. And in the last equality we used that the event $\{u < \frac{\tau(t-1)}{X_\infty}\}$ includes the event $\{u \leq \frac{X_\infty - \tau}{X_\infty}\}$ since $\tau t < 1 \leq X_\infty$. Finally,

$$\mathbb{P}\{X_\infty < t\} = 2\mathbb{P}\left\{u < \frac{\tau(t-1)}{X_\infty}\right\} = 2\mathbb{E}\left(\mathbb{P}\left\{\frac{\tau(t-1)}{X_\infty} > u\right\} \mid X_\infty\right)$$

$$= 2\mathbb{E}\left(\frac{\tau(t-1)}{X_\infty}\right) = 2\tau\mathbb{E}\frac{1}{X_\infty}(t-1). \tag{80}$$

And we also have a special state $X_\infty = 2$:

$$\begin{aligned} \mathbb{P}\{X_\infty = 2\} &= \mathbb{P}\left\{1 + 1 = 2, \frac{uX_\infty}{\tau} \geq 1, \frac{(1-u)X_\infty}{\tau} \geq 1\right\} \\ &= \mathbb{P}\left\{u \geq \frac{\tau}{X_\infty}, u \leq 1 - \frac{\tau}{X_\infty}\right\} = 1 - 2\tau\mathbb{E}\frac{1}{X_\infty}. \end{aligned} \tag{81}$$

From (80) and (81) we see that X_∞ is a combination of the uniform distribution on $[1, 2)$ and an atom at point 2. We only need to compute a constant $\mathbb{E}\frac{1}{X_\infty}$.

$$\begin{aligned} \mathbb{E}\frac{1}{X_\infty} &= \frac{1}{2}\mathbb{P}\{X_\infty = 2\} + \int_1^2 \frac{1}{t}d\mathbb{P}\{X_\infty < t\} \\ &= \frac{1}{2}\left(1 - 2\tau\mathbb{E}\frac{1}{X_\infty}\right) + \int_1^2 \frac{2\tau\mathbb{E}\frac{1}{X_\infty}}{t}dt \\ &= \frac{1}{2}\left(1 - 2\tau\mathbb{E}\frac{1}{X_\infty}\right) + 2\tau\mathbb{E}\frac{1}{X_\infty}(\log 2 - \log 1). \end{aligned} \tag{82}$$

We solve (82) for an unknown $\mathbb{E}\frac{1}{X_\infty}$ and obtain that

$$\mathbb{E}\frac{1}{X_\infty} = \frac{1}{2\tau(1 - 2\log 2 + 1/\tau)}. \tag{83}$$

On Figure 1 the reader can see simultaneous plots of just obtained theoretical values of $F(t)$ and statistical values. We refer to statistical values as an empirical cumulative distribution function obtained from values of $X_h^{(i)}$ taken in 10000 consecutive iterations of the process (5).

From now on we assume that $\tau > 1/2$. We begin as in previous case but write all possibilities for minimums separately.

The situation $X_\infty = 2$ is possible only when both minimums are 1.

$$\begin{aligned} \mathbb{P}\{X_\infty = 2\} &= \mathbb{P}\left\{1 + 1 = 2, 1 \leq \frac{uX_\infty}{\tau}, 1 \leq \frac{(1-u)X_\infty}{\tau}\right\} \\ &= \mathbb{P}\left\{\frac{\tau}{X_\infty} \leq u \leq \frac{X_\infty - \tau}{X_\infty}, X_\infty \geq 2\tau\right\} \\ &= \int_{[2\tau, 2]} \left(1 - \frac{\tau}{x}\right) dF(x) - \int_{[2\tau, 2]} \frac{\tau}{x} dF(x) \\ &= 1 - F(2\tau) - 2\tau \int_{[2\tau, 2]} \frac{1}{x} dF(x) \\ &= 1 - F(2\tau) - 2\tau \left(\mathbb{E}\frac{1}{X_\infty} - \int_{[1, 2\tau)} \frac{1}{x} dF(x)\right). \end{aligned} \tag{84}$$

When $X_\infty < 2$ there are two possibilities that only one of the minimums in (62) is 1, and both of these situations are symmetrical, so we compute only one of the

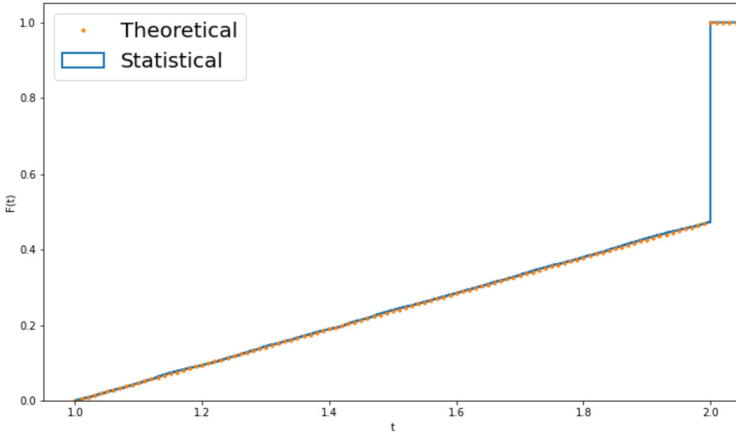


Fig. 1. Plot of theoretical and statistical values of the distribution function $F(t)$ of X_∞ with $\tau = 0.4$

probabilities:

$$\begin{aligned}
 & \mathbb{P} \left\{ \frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1 \right\} \\
 &= \mathbb{P} \left\{ u < \frac{\tau(t-1)}{X_\infty}, X_\infty \geq \tau t \right\} + \mathbb{P} \left\{ u \leq \frac{X_\infty - \tau}{X_\infty}, X_\infty < \tau t \right\} \\
 &= \int_{[\tau t, 2]} \frac{\tau(t-1)}{x} dF(x) + \int_{[1, \tau t]} \left(1 - \frac{\tau}{x}\right) dF(x) \\
 &= \tau(t-1) \left(\mathbb{E} \frac{1}{X_\infty} - \int_{[1, \tau t]} \frac{1}{x} dF(x) \right) + F(\tau t) - \tau \int_{[1, \tau t]} \frac{1}{x} dF(x). \tag{85}
 \end{aligned}$$

And the last possibility of $X_\infty < 2$ is when both minimums are less than 1.

$$\begin{aligned}
 & \mathbb{P} \left\{ \frac{uX_\infty}{\tau} + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} < 1 \right\} \\
 &= \mathbb{P} \left\{ \frac{X_\infty - \tau}{X_\infty} < u < \frac{\tau}{X_\infty}, x < \tau t \right\} = \int_{[1, \tau t]} \frac{\tau}{x} dF(x) - \int_{[1, \tau t]} \left(1 - \frac{\tau}{x}\right) dF(x) \\
 &= 2\tau \int_{[1, \tau t]} \frac{1}{x} dF(x) - F(\tau t). \tag{86}
 \end{aligned}$$

We combine (85) and (86) and get the following for $t \in [1, 2]$:

$$\begin{aligned}
 \mathbb{P}\{X_\infty < t\} &= \mathbb{P} \left\{ \frac{uX_\infty}{\tau} + \frac{(1-u)X_\infty}{\tau} < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} < 1 \right\} \\
 &\quad + 2\mathbb{P} \left\{ \frac{uX_\infty}{\tau} + 1 < t, \frac{uX_\infty}{\tau} < 1, \frac{(1-u)X_\infty}{\tau} \geq 1 \right\} \\
 &= F(\tau t) + 2\tau(t-1) \left(\mathbb{E} \frac{1}{X_\infty} - \int_{[1, \tau t]} \frac{1}{x} dF(x) \right). \tag{87}
 \end{aligned}$$

Now we will show how to retrieve exact values for probability (87). On step 1 we consider that $t \in [1, 1/\tau]$, then $\tau t < 1$ and from (87) one gets:

$$\mathbb{P}\{X_\infty < t\} = 0 + 2\tau(t - 1) \left(\mathbb{E} \frac{1}{X_\infty} - 0 \right) = 2\tau(t - 1) \mathbb{E} \frac{1}{X_\infty}. \tag{88}$$

We also compute the derivative

$$dF(t) = 2\tau \mathbb{E} \frac{1}{X_\infty} dt. \tag{89}$$

We repeat the procedure by computing probabilities and the derivatives for $t \in [1/\tau, 1/\tau^2]$ and so on following intervals by using results from previous intervals. More generally, let us compute an interval on step k when $t \in [1/\tau^{k-1}, 1/\tau^k]$.

We have $\tau t \in [1/\tau^{k-2}, 1/\tau^{k-1}]$ which means that $F(\tau t)$ is a value from previous interval and we can use it. We also have an integral part in (87):

$$\int_{[1, \tau t]} \frac{1}{x} dF(x) = \int_1^{1/\tau} \frac{1}{x} dF(x) + \int_{1/\tau}^{1/\tau^2} \frac{1}{x} dF(x) + \dots + \int_{1/\tau^{k-2}}^{\tau t} \frac{1}{x} dF(x) \tag{90}$$

where the integrals are taken by using computed derivatives on first $k - 1$ intervals respectively.

At some k we have $1/\tau^k > 2$ so we stop computing intervals. We have only one unknown variable $\mathbb{E} \frac{1}{X_\infty}$. We compute this value directly by using computed probabilities and solve an equation for it.

Example 4.5. Let us compute X_∞ when $1/\tau < 2 \leq 1/\tau^2$ ($k = 2$). We also have that $2\tau \leq 1/\tau$.

When $t \leq 1/\tau$, we have

$$\mathbb{P}\{X_\infty < t\} = 2\tau(t - 1) \mathbb{E} \frac{1}{X_\infty}, \quad dF(t) = 2\tau \mathbb{E} \frac{1}{X_\infty} dt. \tag{91}$$

When $t \geq 1/\tau$,

$$\int_{[1, \tau t]} \frac{1}{x} dF(x) = \int_1^{\tau t} \frac{2\tau \mathbb{E} \frac{1}{X_\infty}}{x} dx = 2\tau \mathbb{E} \frac{1}{X_\infty} \log(\tau t). \tag{92}$$

After plugging (91) and (92) into (87) we obtain

$$\begin{aligned} \mathbb{P}\{X_\infty < t\} &= F(\tau t) + 2\tau(t - 1) \left(\mathbb{E} \frac{1}{X_\infty} - 2\tau \mathbb{E} \frac{1}{X_\infty} \log(\tau t) \right) \\ &= 2\tau(\tau t - 1) \mathbb{E} \frac{1}{X_\infty} + 2\tau(t - 1) \mathbb{E} \frac{1}{X_\infty} - 4\tau^2(t - 1) \log(\tau t) \\ &= 2\tau \mathbb{E} \frac{1}{X_\infty} (t + \tau t - 2\tau t \log(\tau t) + 2\tau \log(\tau t) - 2), \end{aligned} \tag{93}$$

and the derivative is

$$dF(t) = 2\tau \mathbb{E} \frac{1}{X_\infty} \left(1 + \tau - 2\tau(1 + \log(\tau t)) + 2\tau \frac{1}{t} \right). \tag{94}$$

Finally, we also return to a special case value $X_\infty = 2$ which was computed at the beginning in (84):

$$\begin{aligned} \mathbb{P}\{X_\infty = 2\} &= 1 - F(2\tau) - 2\tau \left(\mathbb{E} \frac{1}{X_\infty} - \int_{[1, 2\tau)} \frac{1}{x} dF(x) \right) = 1 - \mathbb{P}\{X_\infty < 2\} \\ &= 1 - 2\tau \mathbb{E} \frac{1}{X_\infty} (2 + 2\tau - 4\tau \log(2\tau) + 2\tau \log(2\tau) - 2) \\ &= 1 - 4\tau^2 \mathbb{E} \frac{1}{X_\infty} (1 - \log(2\tau)). \end{aligned} \tag{95}$$

We only need to compute the constant $\mathbb{E} \frac{1}{X_\infty}$:

$$\begin{aligned} \mathbb{E} \frac{1}{X_\infty} &= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E} \frac{1}{X_\infty} (1 - \log(2\tau)) \right) + \int_1^{1/\tau} \frac{2\tau \mathbb{E} \frac{1}{X_\infty}}{x} dx \\ &\quad + \int_{1/\tau}^2 \frac{2\tau \mathbb{E} \frac{1}{X_\infty} (1 + \tau - 2\tau(1 + \log(\tau t)) + 2\tau \frac{1}{t})}{x} dx \\ &= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E} \frac{1}{X_\infty} (1 - \log(2\tau)) \right) + \int_1^{1/\tau} \frac{2\tau \mathbb{E} \frac{1}{X_\infty}}{x} dx \\ &\quad + 2\tau \mathbb{E} \frac{1}{X_\infty} \left(\int_{1/\tau}^2 \frac{1 + \tau - 2\tau}{x} dx + \int_{1/\tau}^2 \frac{-2\tau \log \tau x}{x} dx + \int_{1/\tau}^2 \frac{2\tau}{x^2} dx \right) \\ &= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E} \frac{1}{X_\infty} (1 - \log(2\tau)) \right) + \int_1^{1/\tau} \frac{2\tau \mathbb{E} \frac{1}{X_\infty}}{x} dx + 2\tau \mathbb{E} \frac{1}{X_\infty} * \\ &\quad * \left((1 - \tau)(\log 2 - \log(1/\tau)) - \tau(\log^2(2\tau) - \log^2(\tau/\tau)) - 2\tau \left(\frac{1}{2} - \tau \right) \right) \\ &= \frac{1}{2} \left(1 - 4\tau^2 \mathbb{E} \frac{1}{X_\infty} (1 - \log(2\tau)) \right) - 2\tau \mathbb{E} \frac{1}{X_\infty} \log(\tau) \\ &\quad + 2\tau \mathbb{E} \frac{1}{X_\infty} \left((1 - \tau) \log(2\tau) - \tau \log^2(2\tau) + 2\tau \left(\tau - \frac{1}{2} \right) \right). \end{aligned} \tag{96}$$

By solving this equation for an unknown $\mathbb{E} \frac{1}{X_\infty}$ we obtain:

$$\mathbb{E} \frac{1}{X_\infty} = \frac{1}{2(1 - 4\tau^3 + 4\tau^2 + 2\tau^2 \log^2(2\tau) + 2\tau \log(\tau) - 2\tau \log(2\tau))}. \tag{97}$$

As the end of the example, Figure 2 shows plots of theoretical and statistical values of $F(t)$ with τ lying in the studied range.

Finally, without computing theoretical values of the distribution function, we want to show how it further evolves with increasing τ by showing only its statistical values on Figure 3 with $\tau = 0.9$.

Acknowledgments

I would like to thank my supervisor, Prof. Alexander Marynych, for his dedicated guidance and support throughout this research. I am grateful for his prompt and

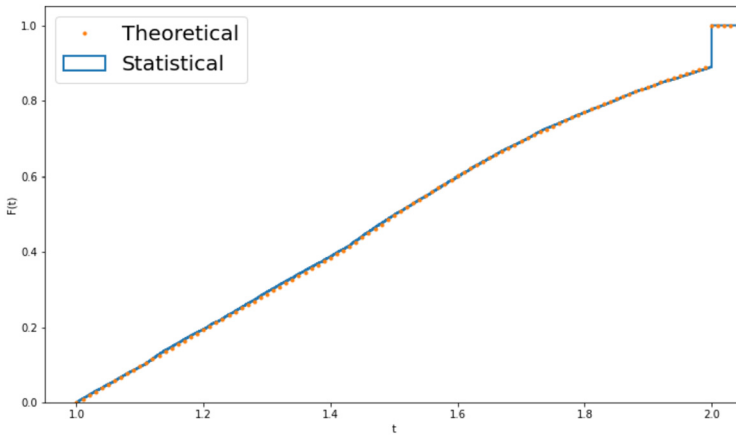


Fig. 2. Plot of theoretical and statistical values of the distribution function $F(t)$ of X_∞ with $\tau = 0.7$

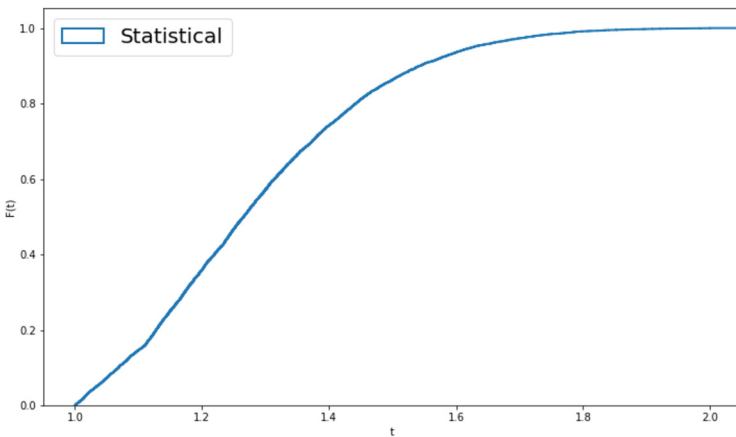


Fig. 3. Plot of statistical values of the distribution function $F(t)$ of X_∞ with $\tau = 0.9$

friendly responses to my questions and queries, as well as for his genuine care about my work. I also thank the anonymous referee for useful suggestions which improved the presentation of my results.

Funding

The author was supported by the National Research Foundation of Ukraine (project 2020.02/0014 “Asymptotic regimes of perturbed random walks: on the edge of modern and classical probability”).

References

- [1] Ambrus, G., Kevei, P., Vigh, V.: The diminishing segment process. *Stat. Probab. Lett.* **82**(1), 191–195 (2012). [MR2863042](#). <https://doi.org/10.1016/j.spl.2011.09.016>
- [2] Bavaud, F.: Adjoint transform, overconvexity and sets of constant width. *Trans. Am. Math. Soc.* **333**(1), 315–324 (1992). [MR1132431](#). <https://doi.org/10.2307/2154111>
- [3] Bentley, J.: Multidimensional binary search trees used for associative searching. *Commun. ACM* **18**(9), 509–517 (1975)
- [4] Billingsley, P.: *Convergence of Probability Measures*, Second Edition. John Wiley & Sons, Inc., New York (1999). [MR1700749](#). <https://doi.org/10.1002/9780470316962>
- [5] Cover, T.: Estimation by the nearest neighbor rule. *IEEE Trans. Inf. Theory* **14**(1), 50–55 (1968)
- [6] Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**(1), 21–27 (1967)
- [7] Devroye, L.: The uniform convergence of nearest neighbor regression function estimators and their applications in optimization. *IEEE Trans. Inf. Theory* **24**(2), 142–151 (1978). [MR0518942](#). <https://doi.org/10.1109/tit.1978.1055865>
- [8] Durrett, R.: *Probability Theory and Examples*, Fourth Edition. Cambridge University Press, New York (2010). [MR2722836](#). <https://doi.org/10.1017/CBO9780511779398>
- [9] Fu, A., Chan, P., Cheung, Y., Moon, Y.: Dynamic vp-tree indexing for n-nearest neighbor search given pair-wise distances. *VLDB J.* **9**, 154–173 (2000)
- [10] Kevei, P., Vigh, V.: On the diminishing process of Balint Toth. *Trans. Am. Math. Soc.* **368**(12), 8823–8848 (2016). [MR3551590](#). <https://doi.org/10.1090/tran/6620>
- [11] Kumar, N., Zhang, L., Nayar, S.: What is a good nearest neighbors algorithm for finding similar patches in images. In: *Lecture Notes in Computer Science*, Vol 5303, pp. 364–378. Springer, Berlin (2008)
- [12] Moore, A.: An introductory tutorial on kd-trees. Technical Report Technical Report No. 209, Computer Laboratory, University of Cambridge, Carnegie Mellon University, Pittsburgh, PA (1991)
- [13] Narasimhulu, Y., Suthar, A., Pasunuri, R., Venkaiah, V.: Ckd-tree: An improved kd-tree construction algorithm. In: *International Semantic Intelligence Conference 2021*, pp. 211–218 (2021)
- [14] Omohundro, S.: Five balltree construction algorithms. Technical Report Technical Report 89-063 (1989)
- [15] Papernot, N., McDaniel, P.: Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning. *CoRR* [1803.04765](#) (2018)
- [16] Shishibori, M., Lee, S., Kita, K.: An improved method to select candidates on metric index vp-tree. In: *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, vol. 1, pp. 306–311 (2011)
- [17] Yianilos, P.: Data structures and algorithms for nearest neighbor search in general metric spaces. In: *Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 311–321 (1993). [MR1213243](#)