# RANDOM HYPERPLANE SEARCH TREES IN HIGH DIMENSIONS *

*Luc Devroye*† *and James King*‡

Abstract. Given a set $S$ of $n \geq d$ points in general position in $\mathbb{R}^d$, a *random hyperplane split* is obtained by sampling $d$ points uniformly at random without replacement from $S$ and splitting based on their affine hull. A *random hyperplane search tree* is a binary space partition tree obtained by recursive application of random hyperplane splits. We investigate the structural distributions of such random trees with a particular focus on the growth of $d$. A *blessing of dimensionality* arises—as $d$ increases, random hyperplane splits more closely resemble perfectly balanced splits; in turn, random hyperplane search trees more closely resemble perfectly balanced binary search trees.

We prove that, for any fixed dimension $d$, a random hyperplane search tree storing $n$ points has height at most $(1 + \mathcal{O}(1/\sqrt{d}))\log_2 n$ and average element depth at most $(1 + \mathcal{O}(1/d))\log_2 n$ with high probability as $n \to \infty$. Further, we show that these bounds are asymptotically optimal with respect to $d$.

## 1 Introduction

Point sets in $\mathbb{R}^d$ can be partitioned recursively by a number of possible trees. The early, and still most popular, choices are the $k$-d tree and the quadtree. Later options include binary space partition (or BSP) trees [32]. The $k$-d tree takes a point from the set and partitions the space into two sets with a hyperplane containing the point that is perpendicular to one of the axes. In a quadtree, the split is into $2^d$ quadrants obtained by shifting the origin to the point in question. A lot of ink has been spilled on the analysis of the shapes of the trees for random point sets—for a summary and mini-survey, see the results of Devroye [10].

In this paper, we focus on deterministic point sets, outside the control of the user, and random partitions that are built on them. For example, in either of the two trees mentioned above, one could choose a splitting point uniformly at random, and make independent choices recursively on the subsets. We assume throughout that points are in general position (no three on a line, no four on a plane, and so forth). In both examples, if the set of data points lies on the moment curve $\{(x, x^2, \ldots, x^d) : x \in \mathbb{R}\}$, then the tree thus obtained is statistically equivalent to a random binary search tree.

Analysis of random tree data structures typically focusses on two functions that quantify the level of balance: the depth (specifically, the mean point depth) and the height (*i.e.*, the maximum point depth). These values are of particular practical importance—when searching for point in the tree, they correspond respectively to the average-case and

*This research was supported by NSERC grant A3456.

†*McGill University*, luc.devroye@gmail.com

‡*D-Wave Systems*, jamie.king@gmail.com

worst-case query times. For a perfectly balanced binary search tree, using $H_n^*$ to denote the height (maximal path distance from leaf to root), we have

$$\lim_{n\to\infty} \frac{H_n^*}{\log_2 n} = 1,$$

which is the best we can hope for when dealing with binary trees.

A random $k$-d tree in any dimension has the same shape as a random binary search tree—notably, the distribution does not depend on the structure of the point set, only its size. We use $H_n$ to denote the height of a tree storing $n$ points. Let $D_n$ denote the depth (path distance to the root) of a randomly selected node. It is known that $D_n/\log n \to 2$ in probability [25, 23, 7]. The limit law for $D_n$ was derived by Devroye [7]: $(D_n - 2\log n)/\sqrt{2\log n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution. Robson [36], Pittel [34], and Devroye [5, 6] showed that $H_n/\log n \to 4.31107\ldots$ in probability. See Mahmoud [26] for more background.

### 1.1 Random hyperplane search trees

For a given set $S$ of $n$ points in general position in $\mathbb{R}^d$, a *random hyperplane search tree* is constructed as follows. If $n \geq d$, it selects at the root level $d$ points uniformly at random without replacement from the $n$ data points, and considers the hyperplane through these points, *i.e.*, their affine hull. These $d$ pivot points are associated with the root node and remain there. The hyperplane splits the remaining $n - d$ points into two sets that are handled recursively and independently; a fair coin flip decides which set goes to the left subtree and which goes to the right subtree. If $n < d$, no splitting is applied, and all $n$ points are associated with the root, which becomes a leaf. This construction guarantees that each internal node holds $d$ data points and each leaf node holds between 0 and $d - 1$ data points.

A key feature of hyperplane search trees is that they are constructed independently of the axes and are therefore robust to affine transformations of the underlying point set. If the set of points contained by a $k$-d tree undergoes a rotation, the $k$-d tree would have to be reconstructed. However, this is not the case for hyperplane search trees (see Figures 1 and 2).

**Applications** Hyperplane search trees have been used since the 1970s in many applications of statistics. For example, Mizoguchi *et al.* [30] highlighted their use in pattern recognition and You and Fu [44] considered their use as *tree classifiers*. In computational geometry, trees based upon partitions of space by means of hyperplanes are ubiquitous. See for example the survey of Edelsbrunner and Van Leeuwen [13], or the work of Haussler and Welzl [18] on simplex range queries. For more examples and references, see our previous paper [11, §2].

Fuchs, Kedem and Naylor [15] introduce the BSP TREES ("binary space partition trees") for use in graphics applications. The space is split in two linear halfspaces; each halfspace may in turn be split by a linear hyperplane, and so forth. If a viewer sits in a

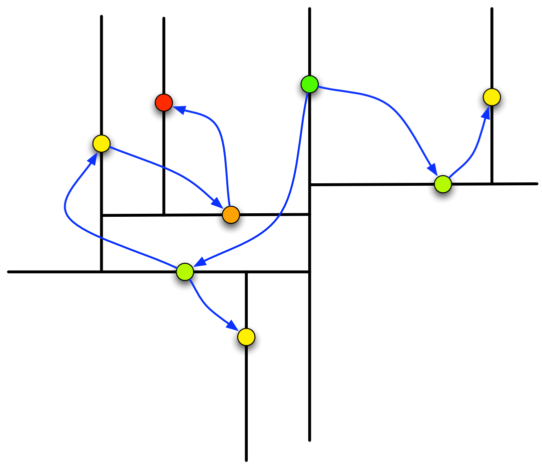

Figure 1: A $k$-d tree in $\mathbb{R}^2$. Rotating the point set could make the divisions non-axis-parallel, thereby invalidating the tree.

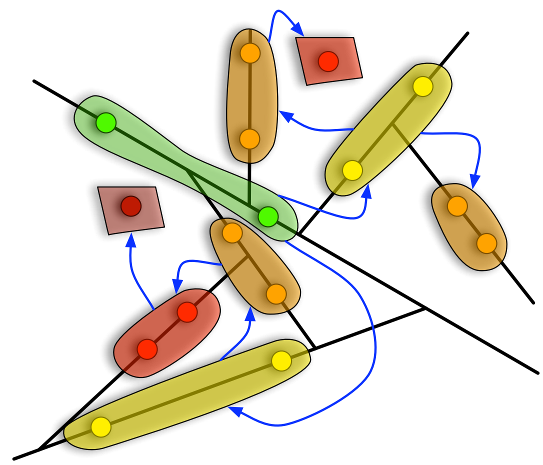

Figure 2: A hyperplane search tree in $\mathbb{R}^2$. If the point set is rotated, the divisions rotate with the points and remain valid.

given polyhedral set in this partition, and wants to project the world onto his/her view plane, the BSP tree aids in establishing the order in which the polyhedral cells must be drawn so as not to cause visibility problems. Basically, one should consider polyhedra in depth-first-search order, where the depth-first-search first visits halfspaces that would not contain the viewer, so that polyhedra are visited from "far" to "near" (this is called the painter's algorithm). For more on the hidden surface elimination with the aid of BSP trees, see Paterson and Yao [32], Samet [38, 37], or Fuchs, Abram and Grant [14]. While BSP trees are not hyperplane trees (because in general, one does not take data points to generate the partition), they are intimately related and indicate interesting applications of hyperplane trees in hidden surface elimination and beam tracing. See also Paterson and Yao [32], Sung and Shirley [39] and Kaplan [21].

### 1.2 Results

Define $\mathcal{S}_{n,d} = \{S : S \subset \mathbb{R}^d, |S| = n, S \text{ is in general linear position}\}$. For a set $S \in \mathcal{S}_{n,d}$ we use $H(S)$ and $D(S)$ to denote the height and depth of a random point of a random hyperplane search tree built on $S$: $H(S)$ and $D(S)$ are random variables. By a trivial coupling argument we have that $H(S)$ stochastically dominates $D(S)$, *i.e.*, $\mathbb{P}\{D(S) \le t\} \ge \mathbb{P}\{H(S) \le t\}$ for any value of $t$. In order to more cleanly express bounds on $H(S)$ and $D(S)$, we define

$$C_H(d) \stackrel{\text{def}}{=} \inf\left\{c \in \mathbb{R} : \lim_{n\to\infty} \max_{S\in\mathcal{S}_{n,d}} \mathbb{P}\left\{\frac{H(S)}{\log n} \le c\right\} = 1\right\}, \tag{1}$$

$$C_D(d) \stackrel{\text{def}}{=} \inf\left\{c \in \mathbb{R} : \lim_{n\to\infty} \max_{S\in\mathcal{S}_{n,d}} \mathbb{P}\left\{\frac{D(S)}{\log n} \le c\right\} = 1\right\}. \tag{2}$$

The phenomenon that we wish to investigate is that uniformly over all sets $S \in \mathcal{S}_{n,d}$, the behavior of $H(S)$ and $D(S)$ is nearly optimal when $d$ is large. It is already known [11] that $C_H(1) = C_H(2) = 4.31107\ldots$ and that $C_H(d) < C_H(1)$ for $d \geq 3$, thus showing that hyperplane search trees outperform random binary search trees or $k$-d trees for all dimensions, with the improvement being strict when $d \geq 3$. The present note makes this more precise, and shows that in fact, $\lim_{d\to\infty} C_H(d) = 1/\log 2$. Thus, by pushing up $d$, we can ensure almost perfectly balanced trees almost all the time, as we have the trivial lower bound $H(S) \geq \log_2 n$. We also derive expressions and bounds on $C_H(d)$ and $C_D(d)$ as we proceed.

It is natural to go beyond this result and ask how quickly random hyperplane search trees become balanced as $d$ increases. We find that the constants corresponding to height and average depth decay at different rates. The main contribution of this paper is proving asymptotically optimal bounds for these rates, stated in the following two theorems:

**Theorem 1.** *We have $C_H(d) = 1/\log 2 + \mathcal{O}\big(d^{-1/2}\big)$. This bound is asymptotically optimal since there exists a function $g_H(d) = 1/\log 2 + \Omega\big(d^{-1/2}\big)$ such that*

$$\lim_{n\to\infty} \max_{S\in\mathcal{S}_{n,d}} \mathbb{P}\left\{ \frac{H(S)}{\log n} \geq g_H(d) \right\} = 1.$$

**Theorem 2.** *We have $C_D(d) = 1/\log 2 + \mathcal{O}\big(d^{-1}\big)$. This bound is asymptotically optimal since there exists a function $g_D(d) = 1/\log 2 + \Omega\big(d^{-1}\big)$ such that*

$$\lim_{n\to\infty} \max_{S\in\mathcal{S}_{n,d}} \mathbb{P}\left\{ \frac{D(S)}{\log n} \geq g_D(d) \right\} = 1.$$

### 1.3 Outline

In Section 2 we examine random hyperplane search trees built on *moment curve point sets*. These point sets are conjectured to yield the most unbalanced random hyperplane splits. We discuss their connection with median-of-$(2t+1)$ trees and give simple, closed-form asymptotic lower bounds for the constants governing the height and depth of these trees. These provide the tightness parts—the lower bounds—of Theorems 1 and 2. The remainder of the paper is concerned with upper tail bounds for height and depth.

In Section 3 we prove several lemmas regarding the balance of random hyperplane splits. The bounds in these lemmas are not the strongest known, but unlike other bounds they do not rely on any sophisticated machinery from discrete geometry. These simply-obtained bounds are good enough for us to obtain tight asymptotics for the height and nearly-tight asymptotics for the depth.

In Section 5 we consider the height of dominated trees and prove Theorem 1 using our simple lemmas from Section 3.

In Section 4 we introduce two lemmas providing simple and powerful bounds for the analysis of random split trees. The first lemma bounds the logarithmic moment of a class of random variables that often arise in the analysis of random split trees. The second lemma bounds the depth of a random split tree using a *dominating split variable*. We apply these lemmas to obtain an almost-tight depth bound using our simple geometric lemmas from the previous section.

Finally, in Section 7 we introduce a stronger balance lemma proved by Wagner [40] and restate it in the language of this paper. Using this stronger balance lemma, we prove an asymptotically tight depth bound.

## 2 Moment curve point sets and median-of-$(2t+1)$ trees

The tightness parts of Theorems 1 and 2 can be shown for the moment curve data $X = \{x_1, \ldots, x_n\}$ where

$$x_i = (i, i^2, \ldots, i^d), 1 \leq i \leq n.$$

The points on the moment curve are parametrically ordered, and thus we can order them by first coordinate and refer to the points by their index between 1 and $n$.

Analysis of random hyperplane splits on such point sets is quite clean. Choose $d$ integers uniformly at random from $\{1, \ldots, n\}$ without replacement. This yields $d+1$ (possibly empty) intervals into which the other points fall. Number the intervals. One side of the hyperplane corresponds to all odd-numbered intervals, and the other side to the even-numbered ones. If the intervals catch $N_1, N_2, \ldots, N_{d+1}$ points (with sum $n-d$, of course), then one subtree of the root has size $N_1 + N_3 + \cdots$ data points, and the other one $N_2 + N_4 + \cdots$ data points. A statistically equivalent description yielding the same interval sizes uses $n$ i.i.d. uniform $[0,1]$ random variables $U_1, \ldots U_n$. Use $U_1, \ldots, U_d$ to define the $d+1$ uniform spacings of $[0,1]$, which we call $S_1, \ldots, S_{d+1}$. Then "throw" the remaining $n-d$ points into the intervals. The cardinalities are distributed as $(N_1, \ldots, N_{d+1})$, and are multinomial $(n-d, S_1, \ldots, S_{d+1})$. The size of the odd side of the hyperplane thus is distributed as a sum of multinomial components—it is binomial $(n-d, S_1 + S_3 + \cdots)$. It is well-known that uniform spacings are identically distributed and that their distribution is permutation-invariant (see, e.g., Pyke, 1965). Thus, the odd side of the hyperplane is of size distributed as a binomial $(n-d, S_1 + S_2 + \cdots + S_{(d+1)/2})$ if $d$ is odd and as a binomial $(n-d, S_1 + S_2 + \cdots + S_{(d+2)/2})$ if $d$ is even. But $S_1 + S_2 + \cdots + S_k$ is distributed as a beta $(k, d+1-k)$ random variable. Thus, the root split for the moment curve data yields a left subtree that is binomial $(n-d, \text{beta}((d+1)/2, (d+1)/2))$ when $d$ is odd. For $d$ even, we have with equal probability a binomial $(n-d, \text{beta}((d+2)/2, d/2))$ and a binomial $(n-d, \text{beta}(d/2, (d+2)/2))$. One can verify that this is in turn distributed as a binomial $(n-d, \text{beta}(d/2, d/2))$. If we wish to consider the *fraction* of points on one side of the hyperplane as $n \to \infty$ for some fixed $d$, the expression is even cleaner—the limiting distribution is simply $\text{beta}(\lceil d/2 \rceil, \lceil d/2 \rceil)$ (see, *e.g.*, Devroye [8, Lem. 2] or King [22, §5.2]).

This tree is indistinguishable from the fringe-balanced, or median-of-$(2t+1)$ search tree which has been studied quite extensively in the data structure literature. First suggested by Bell [3] and Walker and Wood [42], it is a binary tree constructed on real-valued data. It samples $2t+1$ data points uniformly without replacement from the $n$ data points, where $t$ is an integer. It then chooses the middle (median) element, and partitions the remaining data points into two sets by using this median point. Assuming without loss of

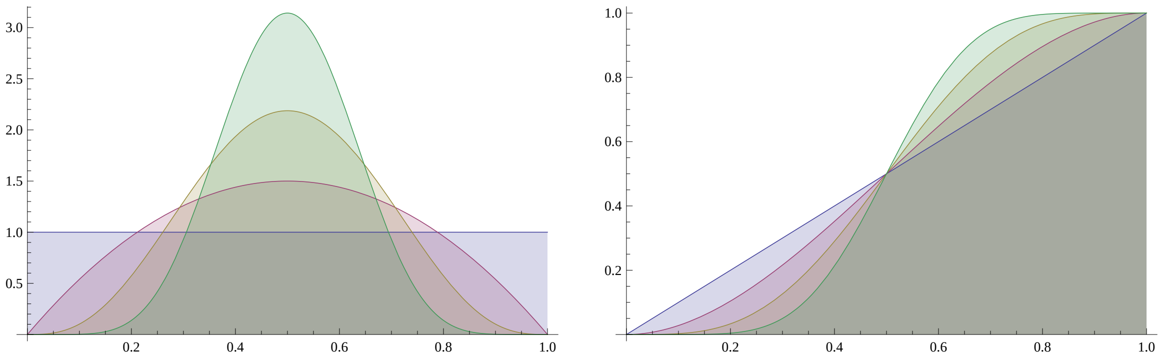


Figure 3: The PDFs (left) and CDFs (right) of the limiting split distribution $\text{beta}(\lceil d/2 \rceil, \lceil d/2 \rceil)$ for $d = 1, 3, 7, 15$. The distribution is uniform when $d = 1$ and becomes more tightly concentrated around $1/2$ as $d$ increases.

generality that the data points are $U_1, \dots, U_n$, as above, we see that the leftmost set in the split is precisely binomial $(n - (2t + 1), \text{beta}(t + 1, t + 1))$. Depending upon the implementation, the $2t$ unused pivot points can also be reused in the partition, thus inflating the subtree sizes by $t$ each. For first-order asymptotics, this is an irrelevant choice. If they are not reused, then the median-of-$(2t + 1)$ tree is distributed as the hyperplane search tree for the moment curve if we take odd $d = 2t + 1$. As we observed above, the moment curve hyperplane search tree for $d = 2t + 2$ is nearly identical, i.e., at least the beta components are of identical parameters. Thus, we will only consider odd $d$.

### 2.1 Lower bounds for the depth and height

The depth $D_n$ of a node uniformly selected from all nodes has been studied by the theory of Markov processes or urn models in a series of papers, notably by Poblete and Munro [35], Aldous et al. [1]. See also Gonnet and Baeza-Yates [16, p. 109] and Devroye [10], where a central limit theorem for $D_n$ can be found. Poblete and Munro [35] showed that

$$\frac{D_n}{\log n} \to \frac{1}{\sum_{i=t+1}^{2t+1} \frac{1}{i+1}} \stackrel{\text{def}}{=} \Lambda(t) \text{ in probability.}$$

Here we give a clean lower bound for $\Lambda(t)$ that proves the second half of Theorem 2.

**Proposition 1.** *As $t \to \infty$,*

$$\Lambda(t) = \frac{1}{\log 2} + \frac{1}{4t \log^2 2} + O\left(\frac{1}{t^2}\right).$$

*Proof.* We know that the $n$th partial sum in the harmonic series is

$$\log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} + \cdots,$$

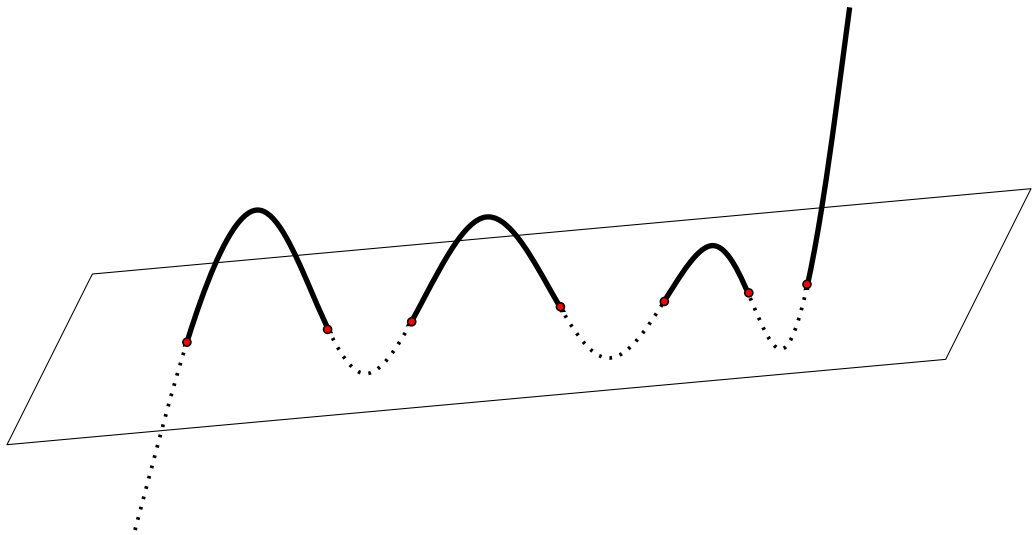

Figure 4: A conceptual visualization of splits caused by choosing random points on the moment curve data set. The data are alternating above and below the hyperplane through the chosen points.

where $\gamma = 0.57721\ldots$ is the Euler-Mascheroni constant. Thus, $\Lambda(t)$ is

$$\frac{1}{\log\left(\frac{2t+2}{t+1}\right) + \frac{1}{4t+2} - \frac{1}{2t+2} + \mathcal{O}\left(\frac{1}{t^2}\right)} = \frac{1}{\log(2) - \frac{1}{4t} + \mathcal{O}\left(\frac{1}{t^2}\right)} = \frac{1}{\log(2)} + \frac{1}{4t\log^2(2)} + \mathcal{O}\left(\frac{1}{t^2}\right).$$

□

The law of large numbers for the height is due to Devroye [9]. We have

$$\frac{H_n}{\log n} \to C(t) \text{ in probability,}$$

where $C(t)$ is the unique solution $c$ greater than $\Lambda(t)$ of the equation

$$\lambda(c) - c\sum_{i=t+1}^{2t+1} \log\left(1 + \frac{\lambda(c)}{i}\right) + c\log 2 = 0,$$

and $\lambda(c)$ is defined by the implicit equation

$$\frac{1}{c} = \sum_{i=t+1}^{2t+1} \frac{1}{\lambda + i}.$$

We have $C(t) \to 1/\log 2$ as $t \to \infty$. A table of numerical values is given in Devroye (1993). For example, for the moment curve in dimensions 1 and 2, the behavior is as for random binary search trees: $D_n/\log n \to 2$ in probability and $H_n/\log n \to 4.31107\ldots$ in probability. In dimension 3, we have a beta$(2,2)$ parameter in the split vector, and obtain

$D_n/\log n \to 12/7$ in probability and $H_n/\log n \to 3.19257\ldots$ in probability. For $d = 2$, this is optimal as shown by Devroye, King and McDiarmid [11]. For $d = 3$, the moment curve yields indeed the worst point configuration, thanks to a result of Welzl [43].

To show the last part of Theorem 1, we need to show that $C(t) \geq 1/\log(2) + c/\sqrt{t}$ for some positive constant $c$ and all $t$ large enough.

**Proposition 2.** *For any constant $c < \sqrt{\log(2)}$ and all $t$ sufficiently large,*

$$C(t) \geq \frac{1}{\log 2} + \frac{c}{\sqrt{t}}.$$

*Proof.* We reparametrize with respect to $t$ as follows. Define $\lambda = \alpha\sqrt{t}$ and $1/c = \log(2) - \beta/\sqrt{t}$. We plug this back into the definitions of $\lambda$ and $c$, and note that it suffices to show that as $t \to \infty$, $\beta$ tends to a positive constant. First note that

$$\begin{aligned}
\sum_{i=t+1}^{2t+1} \frac{1}{\lambda + i} &= \log\left(\frac{\lambda + 2t + 1}{\lambda + t}\right) + \mathcal{O}\left(\frac{1}{t}\right) \\
&= \log(2) + \log\left(\frac{\lambda + 2t + 1}{2\lambda + 2t}\right) + \mathcal{O}\left(\frac{1}{t}\right) \\
&= \log(2) + \log\left(1 - \frac{\lambda - 1}{2\lambda + 2t}\right) + \mathcal{O}\left(\frac{1}{t}\right) \\
&= \log(2) - \frac{\lambda - 1}{2\lambda + 2t} + \mathcal{O}\left(\frac{1}{t}\right) \\
&= \log(2) - \frac{\alpha}{2\sqrt{t}} + \mathcal{O}\left(\frac{1}{t}\right).
\end{aligned}$$

Thus, $|\alpha/2 - \beta| = \mathcal{O}\big(1/\sqrt{t}\big)$. The second equation relating $\lambda$ to $c$ can be rewritten

$$\frac{\lambda}{c\log(2)} + 1 - \frac{1}{\log 2}\sum_{i=t+1}^{2t+1} \log\left(1 + \frac{\lambda}{i}\right) = 0.$$

With the reparametrization, and dividing by $\sqrt{t}$, this yields

$$\alpha\left(1 - \frac{\beta}{\log(2)\sqrt{t}}\right) + \frac{1}{\sqrt{t}} - \frac{1}{\log(2)\sqrt{t}}\sum_{i=t+1}^{2t+1} \log\left(1 + \frac{\alpha\sqrt{t}}{i}\right) = 0.$$

Assuming $\alpha$ remains bounded, the last term is

$$\begin{aligned}
&\frac{1}{\log(2)\sqrt{t}}\sum_{i=t+1}^{2t+1}\frac{\alpha\sqrt{t}}{i}-\frac{1}{\log(2)\sqrt{t}}\sum_{i=t+1}^{2t+1}\frac{\alpha^2 t}{2i^2}+\mathcal{O}\left(\frac{1}{t^2}\right)\\
&=\frac{\alpha}{\log(2)}\left(\log(2)-\frac{1}{4t}+\mathcal{O}\left(\frac{1}{t^2}\right)\right)-\frac{\alpha^2\sqrt{t}}{\log(2)}\sum_{i=t+1}^{2t+1}\frac{1}{2i^2}+\mathcal{O}\left(\frac{1}{t^2}\right)\\
&=\alpha-\frac{\alpha}{4t\log 2}-\frac{\alpha^2}{4\sqrt{t}\log(2)}+\mathcal{O}\left(\frac{1}{t^{3/2}}\right).
\end{aligned}$$

Putting things together, our equation becomes

$$-\frac{\alpha\beta}{\log(2)\sqrt{t}}+\frac{1}{\sqrt{t}}+\frac{\alpha^2}{4\sqrt{t}\log(2)}+\mathcal{O}\left(\frac{1}{t}\right)=0.$$

The main term is $o(1/\sqrt{t})$ if

$$\alpha\beta-\log(2)-\frac{\alpha^2}{4}=o(1).$$

This happens if $\alpha \to \sqrt{\log(16)}$, and thus $\beta \to \sqrt{\log(2)}$. $\square$

## 3 Simple balance lemmas

The following lemma is fundamental.

**Lemma 1** (The small balance lemma). *Consider $d+2$ points in general position in $\mathbb{R}^d$. Let $H$ be the hyperplane through $d$ of them, chosen uniformly at random. Let $A$ be the event that the two remaining points are on the same side of $H$. Then*

$$\mathbb{P}\{A\} \le \frac{1}{2}+\frac{1}{2(d+1)}.$$

*Proof.* Fix $d+2$ points in general position in $\mathbb{R}^d$, denoted by the column vectors $x_1,\ldots,x_{d+2}$, where each $x_i$ contains the $d$ components of the $i$-th point followed by a one. The $(d+1)\times(d+1)$ matrix $B_i$ is defined by

$$B_i=[x_1x_2\cdots x_{i-1}x_{i+1}\cdots x_{d+2}]\,,1\le i\le d+2,$$

i.e., it has all the column vectors except the $i$-th. Let $s_i$ denote the sign of the determinant of $B_i$. This is well-defined as the determinant cannot be zero. The vector $(s_1,\ldots,s_{d+2})$ is all that is needed to conclude the proof. Indeed, we note the following. For a fixed pair $(i,j)$, $i\neq j$ drawn from $\{1,\ldots,d+2\}^2$, let $B_{i,j}$ denote the $(d+1)\times d$ matrix

$$B_{i,j}=[x_1\cdots x_i \text{ and } x_j \text{ missing} \cdots x_{d+2}]\,.$$

Then the $i$-th and $j$-th points are on different sides of the hyperplane that contains the remaining $d$ points if the signs of the determinants of these two matrices are different:

$$[x_i B_{i,j}]\,,[x_j B_{i,j}]\,.$$

Recall that when two adjacent columns of a matrix are swapped, then the determinant changes sign. Repeated swapping to bring $x_i$ or $x_j$ to the front shows that the former matrix has a determinant with sign

$$\begin{cases} s_j(-1)^{i-1} & \text{if } i<j, \\ s_j(-1)^{i} & \text{if } i>j. \end{cases}$$

By symmetry, the sign of the determinant of the latter matrix is

$$\begin{cases} s_i(-1)^{j-1} & \text{if } j<i, \\ s_i(-1)^{j} & \text{if } j>i. \end{cases}$$

The product of the signs is

$$s_i s_j (-1)^{i+j-1}$$

in all cases. Partition $\{1,\ldots,d+2\}$ into four sets,

$$\begin{aligned} A_+ &= \{i : s_i>0, i \text{ even}\}, \\ A_- &= \{i : s_i<0, i \text{ even}\}, \\ B_+ &= \{i : s_i>0, i \text{ odd}\}, \\ B_- &= \{i : s_i<0, i \text{ odd}\}. \end{aligned}$$

Then the number of pairs of points $(i,j)$ with $1 \le i < j \le d+2$ that are on the same side of the hyperplane determined by the $d$ other points is given by

$$\begin{aligned} & \sum_{(i,j):1\le i<j\le d+2} \left( \mathbb{1}_{[i+j \text{ odd}]} \mathbb{1}_{[s_i=s_j]} + \mathbb{1}_{[i+j \text{ even}]} \mathbb{1}_{[s_i\neq s_j]} \right) \\ = \;& |A_+||B_+| + |A_-||B_-| + |A_+||A_-| + |B_+||B_-| \\ = \;& (|A_+|+|B_-|)(|A_-|+|B_+|) \\ = \;& (|A_+|+|B_-|)(d+2-(|A_+|+|B_-|)) \\ \le \;& \begin{cases} \frac{(d+2)(d+2)}{4} & \text{if } d \text{ is even,} \\ \frac{(d+1)(d+3)}{4} & \text{if } d \text{ is odd.} \end{cases} \end{aligned}$$

Thus, the probability that a uniformly randomly selected pair $(i,j)$ ends up at the same side of the hyperplane is not more than the previously computed sum divided by $\binom{d+2}{2}$. This is, for $d$ even, not more than

$$\frac{d+2}{2(d+1)} = \frac{1}{2} + \frac{1}{2(d+1)}.$$

For $d$ odd, it is not more than

$$\frac{d+3}{2(d+2)} = \frac{1}{2} + \frac{1}{2(d+2)}.$$

□

REMARK. For a polytope $P$ of $\mathbb{R}^d$ with $n$ vertices, let $f_k(P)$ denote the number of $k$-faces. A special place is occupied by $\mathcal{C}_{n,d}$, the cyclic polytope in $\mathbb{R}^d$ having $n$ vertices. As a canonical example of such a polytope we can consider the convex hull of the points $\{(t^1, t^2, \ldots, t^d) : t = 1, 2, \ldots, n\}$. McMullen's Upper Bound Theorem (McMullen, 1970; McMullen and Shephard, 1971) [28, 29] states that for all $1 \le k \le d-1$,

$$\max_P f_k(P) = f_k(\mathcal{C}_{n,d}).$$

For more on this, and alternate proofs, see, *e.g.*, Mulmuley (1994) [31], Ziegler (1995) [45], or Kalai (1997) [20]. Exact expression are well-known for $f_k(\mathcal{C}_{n,d})$. The one that is of most interest to us is

$$f_{d-1}(\mathcal{C}_{n,d}) = \binom{n - \lfloor \frac{d+1}{2} \rfloor}{n-d} + \binom{n - \lfloor \frac{d+2}{2} \rfloor}{n-d}.$$

This counts the number of full $(d-1)$-dimensional faces (*i.e.*, facets) of $\mathcal{C}_{n,d}$. For example, when $n = d+2$, one can readily verify these formulas:

$$f_{d-1}(\mathcal{C}_{n,d}) = \begin{cases} \frac{(d+2)^2}{4} & \text{when } d \text{ is even ,} \\ \frac{(d+1)(d+3)}{4} & \text{when } d \text{ is odd.} \end{cases}$$

We also note (see, *e.g.*, Grünbaum (2003) [17]) that if we are given $n = d+2$ points in convex position and in general position, then their convex hull $P$ is a simplicial polytope with $n = d+2$ vertices. Such a polytope must be combinatorially equivalent to $\mathcal{C}_{n,d}$: in particular, $f_k(P) = f_k(\mathcal{C}_{n,d})$ for all $1 \le k \le d-1$. The lemma above relates to these results in a general setting.

REMARK. One of the referees of this paper pointed out a shorter and simpler proof of Lemma 1 that depends on McMullen's Upper Bound Theorem.

*Alternate proof of Lemma 1.* For a set $S$ of $d+2$ points in general position in $\mathbb{R}^d$, there are only three types of $k$-facets, *i.e.*, those for $k \in \{0, 1, 2\}$. The corresponding numbers $e_0, e_1, e_2$ satisfy $e_0 + e_1 + e_2 = 2\binom{d+2}{d} = 2\binom{d+2}{2}$. Moreover, $e_0 = e_2$ equals the number $f_{d-1}$ of facets of the convex hull of $S$. The probability to be upper bounded in Lemma 1 is $\frac{e_0+e_2}{2\binom{d+2}{d}} = \frac{f_{d-1}}{\binom{d+2}{2}}$. By McMullens upper bound theorem, for any set of n points in $\mathbb{R}^d$, the $f_{d-1}$ is maximized if the points are the vertices of a neighbourly polytope (*e.g.*, if the points are on the moment curve), so to prove Lemma 1, it is enough to check the bound for points on the moment curve. For these, it follows easily by just plugging in the known formula for the number of facets. □

Even though it is very simple, we already note that hyperplane splits in sets as small as $d+2$ are roughly balanced for large $d$. Next, we derive an inequality for hyperplane splits for general $n > d$.

**Lemma 2** (The balance lemma)**.** *Consider $n \ge d+1$ points in general position in $\mathbb{R}^d$. Let $H$ be the hyperplane through $d$ of them, chosen uniformly at random. This splits the* *remaining* $n-d$ *points into two sets,* $S$ *and* $S'$*. Let* $N = |S|\xi + |S'|(1-\xi)$*, where* $\xi \in \{0,1\}$ *is* Bernoulli(1/2)*. Then, for* $x \geq 0$*,*

$$\mathbb{P}\left\{N \geq \frac{n-d}{2} + x\right\} \;\leq\; \frac{\frac{n-d}{4} + \frac{(n-d)^2}{4(d+1)}}{\frac{n-d}{4} + \frac{(n-d)^2}{4(d+1)} + x^2}.$$

*Proof.* When $n = d+1$, then the upper bound is more than $1/2$ when $x \leq (n-d)/2$. For $x > (n-d)/2$, the left-hand-side is zero. So assume $n \geq d+2$. Let $H$ denote the random set of $d$ points (instead of the hyperplane that passes through them). For $x_i \notin H$, let $A(x_i, H)$ denote the event that $x_i$ is at the same side of $H$ as the origin (or any other arbitrary fixed point in general position with the others). Set $Y(X_i, H) = 1$ if $A(x_i, H)$ is true and $Y(X_i, H) = -1$ otherwise. Thus,

$$N \;=\; \xi \sum_{i:x_i \notin H} \frac{Y(X_i,H)+1}{2} + (1-\xi)\frac{-Y(X_i,H)+1}{2} \;=\; \frac{n-d}{2} + (\xi - 1/2) \sum_{i:x_i \notin H} Y(X_i, H).$$

Thus, $\mathbb{E}\{N\} = (n-d)/2$, and

$$\begin{aligned}
\mathbf{Var}\{N\} &= \frac{1}{4}\mathbb{E}\left\{\left(\sum_{i:x_i\notin H} Y(X_i,H)\right)^2\right\} \\
&= \frac{1}{4}\mathbb{E}\left\{\sum_{i:x_i\notin H} Y^2(X_i,H)\right\} + \frac{1}{4}\mathbb{E}\left\{\sum_{i\neq j: x_i\notin H, x_j\notin H} Y(x_i,H)Y(x_j,H)\right\} \\
&= \frac{n-d}{4} + \frac{(n-d)(n-d-1)}{4}\mathbb{E}\{Y(x_Z,H)Y(x_W,H)\},
\end{aligned}$$

where $W, Z$ are randomly drawn without replacement from $\{x_1, \ldots, x_n\} \setminus H$. Continuing,

$$\mathbf{Var}\{N\} \;=\; \frac{n-d}{4} + \frac{(n-d)(n-d-1)}{4}\Big(2\cdot\mathbb{P}\{x_Z, x_W \text{ are on same side of } H\} - 1\Big).$$

Now, after first conditioning on the set $H \cup \{x_Z, x_W\}$, which has cardinality $d+2$, Lemma 1 gives us

$$\begin{aligned}
\mathbf{Var}\{N\} &\leq \frac{n-d}{4} + \frac{(n-d)(n-d-1)}{4(d+1)} \\
&\leq \frac{n-d}{4} + \frac{(n-d)^2}{4(d+1)}.
\end{aligned}$$

By the Chebyshev-Cantelli inequality, we have

$$\mathbb{P}\left\{N \geq \frac{n-d}{2} + x\right\} \leq \frac{\mathbf{Var}\{N\}}{\mathbf{Var}\{N\} + x^2}.$$

Plugging in the upper bound on $\mathbf{Var}\{N\}$ gives the result. □

It is convenient to have a simpler bound than that of Lemma 2, in which the sample size $n$ is removed. For example, this suffices for our main result:

**Lemma 3** (The simplified balance lemma)**.** *With notation from Lemma 2, for $x > 1/2$ we have*

$$\mathbb{P}\left\{\frac{N}{n} \geq x\right\} \leq \min\left(\frac{1}{2}, \frac{1}{1+4(d+1)(x-1/2)^2}\right).$$

*Proof.* The 1/2 bound follows from the symmetry in the definition of $N$. We begin by formally replacing $x$ in Lemma 2 by $n(x-1/2)+d/2 = (n-d)(x-1/2)+dx$, and noting that this is $\geq (n-d)(x-1/2)$:

$$\begin{aligned}
\mathbb{P}\{N \geq nx\} &\leq \frac{\frac{n-d}{4} + \frac{(n-d)^2}{4(d+1)}}{\frac{n-d}{4} + \frac{(n-d)^2}{4(d+1)} + (n-d)^2(x-1/2)^2} \\
&= \frac{\frac{1}{4} + \frac{n-d}{4(d+1)}}{\frac{1}{4} + \frac{n-d}{4(d+1)} + (n-d)(x-1/2)^2} \\
&= \frac{\frac{n+1}{4(d+1)}}{\frac{n+1}{4(d+1)} + (n-d)(x-1/2)^2} \\
&= \frac{n+1}{n+1+4(d+1)(n-d)(x-1/2)^2} \\
&\leq \frac{n-d}{n-d+4(d+1)(n-d)(x-1/2)^2} \\
&= \frac{1}{1+4(d+1)(x-1/2)^2}.
\end{aligned}$$

□

## 4 Dominated binary trees

We consider the following general set-up. A tree with $1 \leq n \leq d$ data points is not split and consists of a single node, the root, which "holds" all data points. if $n > d$, the root is split in some manner, resulting in left and right subtree sizes $L, R$, satisfying $L+R = n-d$, and $(L/n, R/n)$ stochastically dominated by $(Z, 1-Z)$, where $Z \in [0,1]$ is a given random variable symmetric about 1/2. By stochastic domination, we mean that

$$\mathbb{P}\{\max(L/n, R/n) \geq x\} \leq \mathbb{P}\{\max(Z, 1-Z)n \geq x\} \ \ (x \geq 0).$$

Define

$$Z^* = \max(Z, 1-Z).$$

From Marshall and Olkin [27], we recall that for any convex function $\psi$,

$$\mathbb{E}\{\psi(L/n) + \psi(R/n)\} \leq \mathbb{E}\{\psi(Z) + \psi(1-Z)\} = 2\mathbb{E}\{\psi(Z)\}.$$

This splitting property is recursively applied to each subtree, and given a subtree size (like $L$), and given the data points that are in the subtree, we require the inequality uniformly over all point sets. To save space, we say that we have a tree dominated by $Z$. Let us give two examples.

**Example 1.** In the random binary search tree, where $d = 1$, we know that $L \stackrel{\mathcal{L}}{=} R \stackrel{\mathcal{L}}{=} \lfloor nU \rfloor$, where $U$ is uniform $[0,1]$. It is trivial to show that $(L/n, R/n)$ is stochastically dominated by $(U, 1-U)$. Thus the tree is dominated by $Z = U$. □

**Example 2.** The hyperplane search tree in $\mathbb{R}^d$. Let the largest of the two subtrees of the root have size $N$. By the union bound and Lemma 3, for $x > 1/2$,

$$\mathbb{P}\left\{\frac{N}{n} \geq x\right\} \leq \frac{2}{1 + 4(d+1)(x - 1/2)^2} < \frac{1}{2(d+1)(x-1/2)^2}.$$

Let $W \in [1/2, 1]$ be a random variable with distribution function given by

$$\mathbb{P}\{W \geq x\} = \begin{cases} \frac{1}{2(d+1)(x-1/2)^2} & \text{if } x \in \left[1/2 + \sqrt{1/2(d+1)}, 1\right], \\ 0 & \text{if } x > 1. \end{cases}$$

That is, $W$ is supported on $\left[1/2 + \sqrt{1/2(d+1)}, 1\right]$ and has an atom of weight $2/(d+1)$ at 1. It takes a moment to verify that

$$W \stackrel{\mathcal{L}}{=} \frac{1}{2} + \min\left(\frac{1}{2}, \sqrt{\frac{1}{2(d+1)U}}\right).$$

Thus, the $Z$ in the preceding discussion can be taken

$$Z = \frac{1}{2} + \sigma \min\left(\frac{1}{2}, \sqrt{\frac{1}{2(d+1)U}}\right),$$

where $\sigma \in \{-1, +1\}$ is a random equiprobable sign. We will see a stronger domination result for hyperplane search trees further on. □

## 5 Height of dominated trees and a proof of Theorem 1

**Lemma 4.** *Consider a tree dominated by $Z$ and define $Z^* = \max(Z, 1-Z)$. For constant $\gamma > 0$, if*

$$\inf_{\lambda > 0} e^{\lambda} \left(\mathbb{E}\left\{2Z^{*\lambda}\right\}\right)^{\gamma} < 1,$$

*then*

$$\lim_{n\to\infty} \mathbb{P}\{H_n > \lceil \gamma \log n \rceil\} = 0.$$

*Proof.* Let $t = \lceil \gamma \log n \rceil$. By domination, we know that both subtrees of the root are stochastically not larger than $nZ^*$. By repeating this observation as we descend away from the root following any path of length $t$, we deduce that the size of the subtree at that node is stochastically not larger than

$$n \prod_{i=1}^{t} Z_i^*,$$

where $Z_1^*, Z_2^*, \ldots$ is an i.i.d. sequence distributed as $Z^*$. Therefore, by the union bound, and the fact that we have a binary tree, using Markov's inequality, and a constant $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\{H_n > t\} &\leq 2^t \mathbb{P}\left\{ n \prod_{i=1}^{t} Z_i^* > d \right\} \\ &\leq 2^t \left(\frac{n}{d+1}\right)^{\lambda} \mathbb{E}\left\{ \left( \prod_{i=1}^{t} Z_i^* \right)^{\lambda} \right\} \\ &= \left(\frac{n}{d+1}\right)^{\lambda} \left( 2\mathbb{E}\left\{ Z^{*\lambda} \right\} \right)^{t}.\end{aligned}$$

The upper bound is not more than

$$\left[ e^{\lambda} \left( \mathbb{E}\left\{ 2Z^{*\lambda} \right\} \right)^{\gamma} \right]^{\log n},$$

which tends to zero if

$$e^{\lambda} \left( \mathbb{E}\left\{ 2Z^{*\lambda} \right\} \right)^{\gamma} < 1.$$

□

*Proof of Theorem 1.* Let us take

$$Z^* \stackrel{\mathcal{L}}{=} \min\left( \frac{1}{2} + a\sqrt{E+b}, 1 \right) \leq \frac{1}{2} + a\sqrt{E+b},$$

where $E$ is standard exponential and $a, b > 0$. Then

$$\mathbb{E}\left\{ (2Z^*)^{\lambda} \right\} \leq \mathbb{E}\left\{ \exp(2a\lambda\sqrt{E+b}) \right\}.$$

Choose $\lambda = 1/(2a)$, and define $\rho = \mathbb{E}\{\exp(\sqrt{E+b})\}$. Then

$$e^{\lambda} \left( \mathbb{E}\left\{ 2Z^{*\lambda} \right\} \right)^{\gamma} \leq \exp\left( \lambda + \gamma(\log(2\rho) - \lambda \log 2) \right) < 1$$

provided

$$\gamma > \frac{\lambda}{\lambda \log 2 - \log(2\rho)} = \frac{1}{\log 2 - 2a \log(2\rho)}.$$

In particular, if $a = \Theta(1/\sqrt{d})$, then this, along with Lemma 4, would imply Theorem 1. But Example 2 implies that a hyperplane search tree is dominated by precisely such a $Z^*$, with $a = 1/\sqrt{2d}$ and $b = \log 8$. □

## 6 Logarithmic moments and depth of dominated trees

The depth of a random node in a tree dominated by $Z$ is determined by the logarithmic moment

$$\mu = 2\mathbb{E}\{Z\log(1/Z)\} = \mathbb{E}\{Z\log(1/Z) + (1-Z)\log(1/(1-Z))\} = \mathbb{E}\{Y\}$$

where $Y$ is a random variable defined as follows:

$$Y = \begin{cases} \log\left(\frac{1}{Z}\right) & \text{with probability } Z, \\ \log\left(\frac{1}{1-Z}\right) & \text{with probability } 1-Z. \end{cases}$$

Note that since $x\log x$ is bounded on $[0,1]$, $\mu \geq 0$ is bounded. Also, $\mu = 0$ if and only if $Z \in \{0,1\}$, i.e., $Z$ is Bernoulli $(1/2)$ (recalling that $Z$ is symmetric). We first provide two useful general lemmas for computing a bound on the logarithmic moment and obtaining a one-sided law of large numbers for general $Z$.

**Lemma 5.** *For a random variable $Z = 1/2 + \sigma V$, where $V$ is $[0,1/2]$-valued, and $\sigma \in \{-1,+1\}$ is a random independent equiprobable sign,*

$$\mu \geq \log 2 - \alpha\mathbb{E}\left\{V^2\right\},$$

*where $\alpha = 2(1+\sqrt{\log 8})^2$.*

*Proof.* Note that

$$-\mu = \mathbb{E}\left\{\left(\tfrac{1}{2}+V\right)\log\left(\tfrac{1}{2}+V\right) + \left(\tfrac{1}{2}-V\right)\log\left(\tfrac{1}{2}-V\right)\right\} = -\log 2 + \mathbb{E}\{f(2V)\},$$

where

$$f(v) \stackrel{\text{def}}{=} \frac{(1+v)\log(1+v)+(1-v)\log(1-v)}{2} \quad (0 \leq v \leq 1).$$

We check that $f(0) = f'(0) = 0$, $f''(v) = 1/(1-v)^2 \geq 0$, so $f$ is convex and increasing to $f(1) = \log 2$. On $[0,b]$, with $b < 1$, we have

$$f(v) \leq \frac{1}{(1-b)^2} \times \frac{v^2}{2}.$$

On $[b,1]$, we have $f(v) \leq \log(2) \leq (v/b)^2 \log 2$. Combining this and choosing $b = \sqrt{\log 8}/(1+\sqrt{\log 8})$, we see that

$$f(v) \leq \frac{1}{4}\alpha v^2 \quad (0 \leq v \leq 1).$$

□

**Lemma 6.** *In a random binary tree dominated by $Z$, having logarithmic moment $\mu > 0$, we have for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{D_n}{\log n} \geq \frac{1}{\mu} + \epsilon\right\} = 0.$$

*Proof.* Let us begin with a small observation. Let $\lambda > 0$ be a parameter and let $X \le 0$ be a nonpositive random variable. Then

$$\lim_{\lambda \downarrow 0} \mathbb{E}\left\{ \frac{e^{\lambda X} - 1}{\lambda} \right\} = \mathbb{E}\{X\}\,.$$

This is best seen by noting that $(e^{\lambda x} - 1)/\lambda \ge x$, which provides a lower bound. Since $(e^{\lambda X} - 1)/\lambda \le 0$, we have by Fatou's lemma,

$$\limsup_{\lambda \downarrow 0} \mathbb{E}\left\{ \frac{e^{\lambda X} - 1}{\lambda} \right\} \le \mathbb{E}\left\{ \limsup_{\lambda \downarrow 0} \frac{e^{\lambda X} - 1}{\lambda} \right\} = \mathbb{E}\{X\}\,.$$

Thus, as $\lambda \downarrow 0$,

$$\varphi(\lambda) \stackrel{\text{def}}{=} \mathbb{E}\left\{ e^{-\lambda Y} \right\} = 1 - \lambda \mathbb{E}\{Y\} + o(\lambda) = 1 - \lambda\mu + o(\lambda).$$

Let us show by induction on the integers $t$ that for all $\lambda \ge 0$, $n \ge 1$,

$$\mathbb{P}\{D_n \ge t\} \le n^\lambda (\varphi(\lambda))\ (t \ge 0).$$

Assuming this for a moment, then we have with $t = \lceil (1/\mu + \epsilon) \log n \rceil$,

$$\begin{aligned}
\mathbb{P}\{D_n \ge t\} &\le n^\lambda (1 - \lambda\mu + o(\lambda))^t \\
&\le \left[ e^\lambda (1 - \lambda\mu + o(\lambda))^{1/\mu + \epsilon} \right]^{\log n} \\
&= [1 - \lambda\mu\epsilon + o(\lambda)]^{\log n} \\
&= o(1)
\end{aligned}$$

if we choose $\lambda > 0$ small enough but fixed. This would complete the proof.

For the proof by induction, note that for $t = 0$, the inequality is trivial. So, we consider a general $t > 0$. Then denoting by $X(L)$ and $X(R)$ the subsets of data points that end up in the left and right subtrees of the root, and by $D_L$ and $D_R$ the depths of random nodes (relative to their subtree roots) of data points randomly selected from $X(L)$ and $X(R)$, respectively, then

$$\begin{aligned}
\mathbb{P}\{D_n \ge t\} &\le \mathbb{E}\left\{ \frac{L}{n} \mathbb{P}\{D_L \ge t-1 | X(L)\} + \frac{R}{n} \mathbb{P}\{D_R \ge t-1 | X(R)\} \right\} \\
&\le \mathbb{E}\left\{ \frac{L}{n} L^\lambda (\varphi(\lambda))^{t-1} + \frac{R}{n} R^\lambda (\varphi(\lambda))^{t-1} \right\} \\
&\le n^\lambda (\varphi(\lambda))^{t-1} \mathbb{E}\left\{ \left(\frac{L}{n}\right)^{\lambda+1} + \left(\frac{R}{n}\right)^{\lambda+1} \right\} \\
&\le n^\lambda (\varphi(\lambda))^{t-1} \mathbb{E}\left\{ Z^{\lambda+1} + (1-Z)^{\lambda+1} \right\} \\
&= n^\lambda (\varphi(\lambda))^{t-1} \mathbb{E}\left\{ e^{-\lambda Y} \right\} \\
&= n^\lambda (\varphi(\lambda))^t.
\end{aligned}$$

□

The inequalities of Section 3 are powerful enough to obtain a depth bound that is almost asymptotically tight. Our proof is of independent interest since it uses only the self-contained Section 3, the previous two lemmas, and simple textbook arguments.

**Proposition 3.** *Consider a hyperplane search tree for a collection of points $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ that are in general position. For fixed $d$, there exists a constant $C(d)$ such that for all $\epsilon > 0$,*

$$\lim_{n\to\infty} \sup_{x_1,\ldots,x_n\in\mathbb{R}^d} \mathbb{P}\{D_n \geq (C(d)+\epsilon)\log_2 n\} = 0.$$

*Furthermore, as $d \to \infty$,*

$$C(d) = 1 + \mathcal{O}\left(\frac{\log(d)}{d}\right).$$

*Proof.* Observe that in the definition of $\mu$, we can take

$$Z = \frac{1}{2} + \sigma V,$$

where

$$V \stackrel{\text{def}}{=} \min\left(\frac{1}{2}, \sqrt{\frac{1}{2(d+1)U}}\right)$$

and $U$ is uniformly distributed on $[0,1]$. By Lemma 5, for this $Z$,

$$\begin{aligned}
\mu &\geq \log 2 - \alpha\mathbb{E}\{V^2\} \\
&= \log 2 - \alpha\mathbb{E}\left\{\min\left(\frac{1}{4}, \frac{1}{2(d+1)U}\right)\right\} \\
&= \log 2 - \frac{\alpha}{4}\mathbb{E}\left\{\min\left(1, \frac{2}{(d+1)U}\right)\right\} \\
&= \log 2 - \frac{\alpha}{2(d+1)}\left(1 + \log\left(\frac{d+1}{2}\right)\right).
\end{aligned}$$

Combining Lemma 6 with this then completes the proof. □

## 7 Stronger bounds on $(\leq k)$-facets and a proof of Theorem 2

Analysis of random hyperplane splits is directly related to the problem of counting $k$-facets in discrete geometry. For a set of $n$ points in general position in $\mathbb{R}^d$, a subset of $d$ points, along with an orientation, defines an oriented hyperplane with an associated positive open halfspace. If this halfspace contains exactly $k$ of the remaining $n-d$ points, we say that the oriented set of $d$ points is a $k$-facet. Thus each subset of $d$ points defines, for some $0 \leq k \leq \lfloor (n-d)/2 \rfloor$, a $k$-facet with one orientation and an $(n-d-k)$-facet with the other orientation. A $(\leq k)$-facet is simply a $j$-facet for some $j \leq k$. Knowing the probability mass function of $N$ is equivalent to knowing the number of $(\leq k)$-facets for every $0 \leq k < \lfloor (n-d)/2 \rfloor$. For a thorough treatment of $k$-facets we direct the reader to Wagner's 2008 survey [41].

A significant open conjecture in discrete geometry is the Spherical Generalized Upper Bound Conjecture, or SGUBC. Forms of this conjecture were proposed independently by Eckhoff [12], Linhart [24], and Welzl [43]. Wagner [40, Conjecture 1.2] proposes the conjecture in full generality. Here we state a slightly weaker form of the conjecture in the language of this paper, which would be implied by SGUBC.

**Conjecture 1.** *For a set of $n \geq d$ points in general position in $\mathbb{R}^d$, define random variable $N$ as the number of points on the larger side of a random hyperplane split. Define $N^*$ as the analogous random variable for the larger side of a random hyperplane split for the moment curve data in $\mathbb{R}^d$. Then, for any $x > (n-d)/2$,*

$$\mathbb{P}\{N \geq x\} \leq \mathbb{P}\{N^* \geq x\}\,.$$

The SGUBC conjecture is trivially true for $d = 1$. For $d = 2$, it was proved by Peck [33] and Alon and Győri [2]. Welzl [43] proved Conjecture 1 for $d = 3$. Inequalities for the far right tail of $N$ were obtained by Clarkson and Shor [4]. For general $d \geq 1$, Wagner [40] proved a relaxed form of a conjecture closely related to the SGUBC that implies Lemma 7 below.

**Lemma 7** (Wagner)**.** *With $N$ and $N^*$ defined as above, for any $x > (n-d)/2$,*

$$\mathbb{P}\{N \geq x\} \leq 4 \cdot \mathbb{P}\{N^* \geq x\}\,.$$

Clarkson and Shor [4] proved a somewhat similar result many years earlier, but their bound only holds for the extreme tail of the distribution, *i.e.*, as $x$ approaches 1, and their proof has hidden factors that depend upon $d$. It is therefore insufficient for our purposes. Wagner's exact expressions and bounds, on the other hand, is valid for the entire range of $x$ that concerns us. In order to exploit Wagner's result using the machinery of the previous section, we must first bound $\mathbb{P}\{N^* \geq x\}$ in an appropriate manner.

**Lemma 8.** *For a set of $n$ distinct points on the moment curve in $\mathbb{R}^d$, we have*

$$\mathbb{P}\left\{\frac{N^*}{n} \geq x\right\} \leq 2\exp\left(-2d\left(x-1/2\right)^2\right) \;\; (x \geq \frac{1}{2}).$$

*Proof.* After first fixing $x \geq 1/2$, for the sake of analysis we introduce random variables

$$B \stackrel{\mathcal{L}}{=} \text{beta}\left(\lceil d/2 \rceil, \lceil d/2 \rceil\right),$$

and $\xi_{d,x}$ with a binomial $(d, x)$ distribution. It is known (see Devroye, 1990) that $\mathbb{P}\{N^*/n \geq y\} \leq \mathbb{P}\{\max(B, 1-B) \geq y\}$ for all $y \geq 0$, *i.e.*, $N^*/n$ is stochastically dominated by $\max(B, 1-B)$. It is also known that the c.d.f.'s of $B$ and $\xi_{d,x}$ are duals: for $x \in (0,1)$,

$$\mathbb{P}\{B \geq x\} = \begin{cases} \mathbb{P}\{\xi_{d,x} \leq (d-1)/2\}\,, & \text{if } d \text{ is odd,} \\ \frac{1}{2}\,\mathbb{P}\{\xi_{d,x} \leq (d-2)/2\} + \frac{1}{2}\mathbb{P}\{\xi_{d,x} \leq d/2\}\,, & \text{if } d \text{ is even.} \end{cases}$$

Thus,

$$\begin{aligned}
\mathbb{P}\left\{\frac{N^*}{n} \geq x\right\} &\leq \mathbb{P}\{\max(B, 1-B) \geq x\} \\
&= 2\mathbb{P}\{B \geq x\} \\
&\leq 2\mathbb{P}\{\xi_{d,x} \leq \lfloor d/2 \rfloor\} \\
&\leq 2\exp\left(-\frac{2(dx - \lfloor d/2 \rfloor)^2}{d}\right) \quad \text{(by Hoeffding's inequality [19])} \\
&\leq 2\exp\left(-\frac{2d^2(x-1/2)^2}{d}\right) \\
&= 2\exp\left(-2d\,(x-1/2)^2\right),
\end{aligned}$$

concluding the proof. □

*Proof of Theorem 2.* Let $E$ be a standard exponential random variable, let $\sigma \in \{-1, +1\}$ be a random equiprobable sign, and define

$$V = \min\left(\frac{1}{2}, \sqrt{\frac{E + \log 8}{2d}}\right), Z = \frac{1}{2} + \sigma V.$$

Note that for $1 \geq x \geq 1/2$,

$$\begin{aligned}
\mathbb{P}\left\{\frac{N}{n} \geq x\right\} &\leq \min\left(4\exp\left(-2d\,(x-1/2)^2\right), 1\right) \\
&= \mathbb{P}\left\{\frac{1}{2} + V \geq x\right\} \\
&= \mathbb{P}\{\max(Z, 1-Z) \geq x\},
\end{aligned}$$

and thus the hyperplane search tree is dominated by this $Z$.

The logarithmic moment $\mu$ of $Z$ is easily bounded using Lemma 5:

$$\begin{aligned}
\mu &\geq \log 2 - \alpha\mathbb{E}\{V^2\} \\
&= \log 2 - \alpha\mathbb{E}\left\{\min\left(\frac{1}{4}, \frac{E + \log 8}{2d}\right)\right\} \\
&= \log 2 - \frac{\alpha}{4d}\mathbb{E}\{\min(d, 2E + 2\log 8)\} \\
&\geq \log 2 - \frac{\alpha}{4d}\mathbb{E}\{2E + 2\log 8\} \\
&= \log 2 - \frac{\alpha(1 + \log 8)}{2d}.
\end{aligned}$$

By Lemma 6, Theorem 2 follows. Thus, the sharper estimates for domination that flow from Wagner's inequality give the optimal rate of convergence with respect to $d$. □

## Acknowledgments

We would like to thank the anonymous referees whose input has helped us clarify and improve the paper.

## References

[1] D. Aldous, B. Flannery, and J.L. Palacios. Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains. *Probability in the Engineering and Informational Sciences*, 2(3):293–307, 1988.

[2] N. Alon and E. Győri. The number of small semispaces of a finite set of points in the plane. *Journal of Combinatorial Theory, Series A*, 41(1):154–157, 1986.

[3] C.J. Bell. *An Investigation into the Principles of the Classification and Analysis of Data on an Automatic Digital Computer*. PhD thesis, Leeds University, 1965.

[4] K.L. Clarkson and P.W. Shor. Applications of random sampling in computational geometry, II. *Discrete and Computational Geometry*, 4(1):387–421, 1989.

[5] L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33:489–498, 1986.

[6] L. Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24:277–298, 1987.

[7] L. Devroye. Applications of the theory of records in the study of random trees. *Acta Informatica*, 26:123–130, 1988.

[8] L. Devroye. On the height of random $m$-ary search trees. *Random Structures and Algorithms*, 1(2):191–204, 1990.

[9] L. Devroye. On the expected height of fringe-balanced trees. *Acta Informatica*, 30(5):459–466, 1993.

[10] L. Devroye. Universal limit laws for depths in random trees. *SIAM Journal on Computing*, 28:409–432, 1999.

[11] L. Devroye, J. King, and C. McDiarmid. Random hyperplane search trees. *SIAM Journal on Computing*, 38(6):2411–2425, 2009.

[12] J. Eckhoff. Helly, Radon, and Carathéodory type theorems. *Handbook of Convex Geometry*, pages 389–448, 1993.

[13] H. Edelsbrunner and J. van Leeuwen. Multidimensional data structures and algorithms: a bibliography. Technical report, Technische Universität Graz, 1983.

[14] H. Fuchs, G.D. Abram, and E.D. Grant. Near real-time shaded display of rigid objects. *Computer Graphics*, 17:65–72, 1983.

[15] H. Fuchs, Z.M. Kedem, and B. Naylor. On visible surface generation by a priori tree structures. In *Proceedings SIGGRAPH '80*, pages 124–133, 1980. Published as Computer Graphics, volume 14.

[16] G.H. Gonnet and R. Baeza-Yates. *Handbook of Algorithms and Data Structures*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1991.

[17] B. Grünbaum. *Convex polytopes*. Springer Verlag, 2003.

[18] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[19] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[20] G. Kalai. Linear programming, the simplex algorithm and simple polytopes. *Mathematical Programming*, 79(1):217–233, 1997.

[21] M.R. Kaplan. The uses of spatial coherence in ray tracing. *SIGGRAPH '85 Course Notes*, 11:22–26, 1985.

[22] J. King. *Guarding Problems and Geometric Split Trees*. PhD thesis, McGill University, 2010.

[23] D.E. Knuth. *The Art of Computer Programming. Vol. 3: Sorting and Searching*. Addison-Wesley, 1973.

[24] J. Linhart. The Upper Bound Conjecture for arrangements of halfspaces. *Contributions to Algebra and Geometry*, 35(1):29–35, 1994.

[25] W.C. Lynch. More combinatorial problems on certain trees. *Computer Journal*, 7:299–302, 1965.

[26] H. M. Mahmoud. *Evolution of Random Search Trees*. John Wiley, New York, 1992.

[27] A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York, 1997.

[28] P. McMullen. The maximum numbers of faces of a convex polytope. *Mathematika*, 17:179–184, 1970.

[29] P. McMullen and G.C. Shephard. *Convex Polytopes and the Upper Bound Conjecture*. Cambridge University Press, 1971.

[30] R. Mizoguchi, M. Kizawa, and M. Shimura. Piecewise linear discriminant functions in pattern recognition. *Systems, Computers, and Control*, 8:114–121, 1977.

[31] K. Mulmuley. *Computational Geometry: an Introduction through Randomized Algorithms*. Prentice Hall, 1994.

[32] M.S. Paterson and F.F. Yao. Efficient binary space partitions for hidden surface removal and solid modeling. *Discrete and Computational Geometry*, 5:485–503, 1990.

[33] G.W. Peck. On $k$-sets in the plane. *Discrete Mathematics*, 56(1):73–74, 1985.

[34] B. Pittel. On growing random binary trees. *Journal of Mathematical Analysis and Applications*, 103:461–480, 1984.

[35] P.V. Poblete and J.I. Munro. The analysis of a fringe heuristic for binary search trees. *Journal of Algorithms*, 6(3):336–350, 1985.

[36] J.M. Robson. The height of binary search trees. *Australian Computer Journal*, 11(4):151–153, 1979.

[37] H. Samet. *Applications of Spatial Data Structures.* Addison-Wesley, Reading, MA, 1990.

[38] H. Samet. *The Design and Analysis of Spatial Data Structures.* Addison-Wesley, Reading, MA, 1990.

[39] K. Sung and P. Shirley. Ray tracing with the BSP tree. In D. Kirk, editor, *Graphics Gems III*, pages 271–274. Academic Press, Boston, 1992.

[40] U. Wagner. On a geometric generalization of the upper bound theorem. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 635–645, 2006.

[41] U. Wagner. $k$-sets and $k$-facets. *Contemporary Mathematics*, 453:443, 2008.

[42] A. Walker and D. Wood. Locally balanced binary trees. *The Computer Journal*, 19(4):322, 1976.

[43] E. Welzl. Entering and leaving $j$-facets. *Discrete and Computational Geometry*, 25(3):351–364, 2001.

[44] K. C. You and K. S. Fu. An approach to the design of a linear binary tree classifier. In *Proceedings of the Symposium of Machine Processing of Remotely Sensed Data*, volume Technical Report 3A-10, Purdue University, 1976.

[45] G.M. Ziegler. *Lectures on Polytopes.* Graduate Texts in Mathematics. Springer, New York, 1995.