One-Sided Interval Trees

Svante Janson

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden. (svante.janson@math.uu.se)

Abstract. We give an alternative treatment and extension of some results of Itoh and Mahmoud on one-sided interval trees. The proofs are based on renewal theory, including a case with mixed multiplicative and additive renewals.

1 Introduction

Itoh and Mahmoud [2] have studied some one-sided versions of binary interval trees. These are obtained from full binary interval trees (see Section 2 for definitions) by pruning one of the two subtrees at each node; in other words, we are left with a single path in the binary interval tree.

Five different such trees, defined by different pruning policies, are studied in [2]. Using an analytic method, Itoh and Mahmoud find explicit or implicit expressions for the moment generating function of the size of the tree, and they derive in each case asymptotic normality and asymptotic expressions for the mean and variance of the size.

We will here give an alternative treatment using renewal theory, which enables us to generalize the results. In particular, the results

cover also the incomplete m-ary interval trees studied by Javanian, Mahmoud and Vahidi-Asl [3].

We will also treat the case of one-sided car-parking, for which Itoh and Mahmoud only give a partial result. While the standard version of one-sided interval trees leads to a multiplicative version of renewal sequences, where standard results from renewal theory are directly applicable, the car-parking version leads to a mixed multiplicative and additive (or rather subtractive) renewal sequence, which requires some new arguments.

2 Definitions and results

Consider repeated divisions of an interval I as follows. If I has length ≥ 1, it is divided into two subintervals I_0 and I_1 by a random division point, which we assume is uniformly distributed in I. The same procedure is then applied recursively to each of the subintervals, until all remaining intervals have lengths less than 1. All random choices are independent.

This construction naturally defines a tree, the *binary interval tree*: The nodes are the intervals that appear, with the original interval I as the root. The intervals of lengths ≥ 1 are internal nodes, each having two children (the subintervals that it is split into). The intervals of lengths < 1 are external nodes or leaves; these are the intervals remaining in the final partition of the original interval.

A *one-sided interval tree* is defined by a similar recursive construction, but each time an interval is split, we now choose one of the two resulting subintervals by some (possibly random) procedure. The chosen subinterval is then divided again (if it has length ≥ 1), but the other subinterval is left undivided, regardless of its length, and is thus represented by a leaf in the tree.

The one-sided interval tree thus consists of a single path of internal nodes with attached leaves. A one-sided interval tree with height H thus has H internal and $H + 1$ external nodes; hence the total size is $2H + 1$.

When we split an interval J in the construction of a one-sided interval tree, its two subintervals have lengths $U|J|$ and $(1 - U)|J|$, where U has a uniform distribution on $(0, 1)$. We then choose one of them according to some policy. We consider only policies where the choice depends only on the relative lengths of the two subintervals, i.e. on U only, possibly with further randomization. Let $p(u)$ be

the probability that we choose the first (left) subinterval, given that $U = u$. The chosen interval then has length $R|J|$ where R equals U or $1 - U$, and conditioned on $U = u$, the probabilities are $p(u)$ and $1 - p(u)$, respectively. It follows that R has a probability density function

$$f_R(u) = p(u) + 1 - p(1 - u), \qquad 0 < u < 1. \tag{2.1}$$

Let the original interval have length $x \geq 1$. Since different intervals are split independently, the first chosen subinterval has length $R_1 x$, the second chosen subinterval has length $R_2 R_1 x$, and so on, where $R_1, R_2, \ldots$ are independent copies of the random variable R with the density function (2.1). Remember that we stop when the interval length becomes less than 1.

We state our first theorem for a general process of this type. We let H_x be the (random) height, or equivalently, the number of internal nodes, when we start with an interval of length x. We further let $T_x = 2H_x + 1$ be the total number of nodes. We state most of our results for H_x while Itoh and Mahmoud [2] state most results for T_x (there denoted S_x); we leave the trivial translation between these to the reader. We are interested in asymptotics as $x \to \infty$.

Theorem 2.1. *Let H_x be the height of a one-sided interval tree, for an initial length x, where each time the chosen subinterval of an interval J (with length $|J| \geq 1$) has length distributed as $R|J|$, for some continuous random variable R with $0 < R < 1$, and we stop when $|J| < 1$. Suppose that $X = -\ln(R)$ has finite mean μ and variance σ^2. Then, as $x \to \infty$,*

$$\mathbb{E}\, H_x = \mu^{-1} \ln x + \frac{\sigma^2 + \mu^2}{2\mu^2} + o(1), \tag{2.2}$$

$$\operatorname{Var} H_x = \frac{\sigma^2}{\mu^3} \ln x + o(\ln x), \tag{2.3}$$

and

$$\frac{H_x - \mu^{-1} \ln x}{\sqrt{\ln x}} \xrightarrow{\mathrm{d}} N(0, \sigma^2/\mu^3). \tag{2.4}$$

For the one-sided interval trees considered here, where R has the

density (2.1), the parameters μ and σ^2 are always finite and given by

$$
\begin{aligned}
\mu &= \mathbb{E}(-\ln(R)) \\
&= \int_0^1 (-\ln u)p(u)\,du + \int_0^1 (-\ln(1-u))(1-p(u))\,du, \\
\sigma^2 &= \mathbb{E}(-\ln(R))^2 - \mu^2 \\
&= \int_0^1 (\ln u)^2 p(u)\,du + \int_0^1 (\ln(1-u))^2(1-p(u))\,du - \mu^2.
\end{aligned}
$$

Itoh and Mahmoud [2] also considered a car-parking version. We begin as above with an interval of length x, but now park cars in it. Each car has, for simplicity, length 1. The cars arrive one by one, and each car parks at a random free place. The first car thus leaves two free subintervals of total length $x - 1$, and the process continues recursively in each of the subintervals until all remaining free intervals have lengths less than 1.

Rényi [4] studied the total number of cars parked by this scheme, i.e. the size of the resulting binary interval tree.

Itoh and Mahmoud considered the corresponding one-sided interval tree, where we each time a car is parked continue in one of the two resulting subintervals only, and ignore (or block) the other subinterval. Thus, if we at some stage have an interval J, with $|J| \geq 1$, we park a car at random in it and obtain two subintervals of lengths $U(|J|-1)$ and $(1-U)(|J|-1)$; we then choose one of them, again according to some policy depending on the relative lengths only. The chosen subinterval thus has length $R(|J|-1)$, with R as above.

We let $\tilde{H}_x$ denote the height of the one-sided interval tree in the car-parking version. Since the cars use some of the space (as can be observed in any city), we obtain smaller intervals in the car-parking version, and thus stop earlier (or at the same time). Thus $\tilde{H}_x \leq H_x$. Nevertheless, as shown in Lemma 3.3 below, the difference is small and asymptotically negligible and we obtain the same first-order asymptotics for both versions.

Theorem 2.2. *Let $\tilde{H}_x$ be the height of a one-sided interval tree, for an initial length x, where each time the chosen subinterval of an interval J (with length $|J| \geq 1$) has length distributed as $R(|J|-1)$, for some continuous random variable R with $0 < R < 1$, and we stop when $|J| < 1$. Suppose that $X = -\ln(R)$ has finite mean μ and*

variance σ^2. Then, as $x \to \infty$,

$$\mathbb{E}\, \tilde{H}_x = \mu^{-1} \ln x + O(1), \tag{2.5}$$

$$\mathrm{Var}\, \tilde{H}_x = \frac{\sigma^2}{\mu^3} \ln x + o(\ln x), \tag{2.6}$$

and

$$\frac{\tilde{H}_x - \mu^{-1} \ln x}{\sqrt{\ln x}} \xrightarrow{\mathrm{d}} N(0, \sigma^2/\mu^3). \tag{2.7}$$

Remark 2.3. The constructions of one-sided interval tree above can be generalized in several ways.

One possibility is to choose the division points with some non-uniform distribution. (I.e., we replace U by another random variable in $(0,1)$.) This leads to the same general setup as above, with some random R (although (2.1) no longer holds). If R has a continuous distribution, this is thus covered by the theorems above. In fact, an inspection of the proofs below shows that we do not really need the assumption that R is continuous; the theorems are valid for any random $R \in (0,1)$ with finite μ and σ^2, except that if R is concentrated on a geometric sequence $\{r^n\}$ for some $r < 1$, the constant term in (2.2) has to be replaced by a term periodic in $\ln x$, or more simply by $O(1)$, cf. [1, Theorem II.5.2(ii)].

Remark 2.4. Javanian, Mahmoud and Vahidi-Asl [3] generalized the problem to random m-ary interval trees, where each interval is randomly divided into m subintervals, where $m \geq 2$ is a fixed integer. Again, they consider the one-sided version where after each division, one of the subintervals is selected and we continue recursively by dividing this interval, as long as its length is at least 1, while the other subintervals are left undivided. With the division and selection rules considered in [3], this becomes another instance of the situation in Theorem 2.1, see Example 4.8 below.

Remark 2.5. Another generalization is to consider car-parking with cars of random lengths. If the cars have a constant length $l \neq 1$, a simple scaling shows that Theorem 2.2 still holds. If the lengths are random (and independent) and lie in an interval $[a, b]$, with $0 < a < b < \infty$, the same holds by comparisons with the cases of constant car lengths a and b. More generally, if the lengths are random, bounded below by some positive number, and have a finite expectation, then Theorem 2.2 still holds by a simple modification of the proof below.

3 Proofs

Proof of Theorem 2.1. By the argument before the theorem, the intervals corresponding to the internal nodes have the lengths x, xR_1, xR_1R_2, ..., where $\{R_i\}_1^\infty$ are i.i.d., and we stop when the length becomes less than 1. Hence, recalling that H_x equals the number of internal nodes,

$$H_x = 1 + \max\Big\{n : x\prod_{i=1}^n R_i \geq 1\Big\}. \tag{3.1}$$

We define $X_i = -\ln(R_i)$ and rewrite (3.1) as

$$H_x - 1 = \max\Big\{n : xe^{-\sum_{i=1}^n X_i} \geq 1\Big\} = \max\Big\{n : \sum_{i=1}^n X_i \leq \ln x\Big\}. \tag{3.2}$$

The quantity in (3.2) is the number of renewals in the interval $(0, \ln x]$ of the sequence of partial sums $S_n = \sum_1^n X_i$. Since the variables X_i are i.i.d. and non-negative, the results now follow from classical results in renewal theory, see for example [1, Theorem II.5.2]. $\qquad\square$

Proof of Theorem 2.2. The lengths of the chosen intervals are

$$L_0 = x,$$
$$L_1 = (L_0 - 1)R_1 = (x - 1)R_1 = xR_1 - R_1,$$
$$L_2 = (L_1 - 1)R_2 = (xR_1 - R_1 - 1)R_2 = xR_1R_2 - R_1R_2 - R_2,$$

and in general, when k cars have parked, by induction,

$$L_k = x\prod_{i=1}^k R_i - \sum_{j=1}^k \prod_{i=j}^k R_i = xe^{-S_k} - \sum_{j=1}^k e^{S_{j-1}-S_k}.$$

We continue until L_k is less than 1. Let

$$Y_k := xe^{-S_k},$$
$$Z_k := \sum_{j=1}^k e^{S_{j-1}-S_k} = e^{-S_k}\sum_{j=0}^{k-1} e^{S_j},$$
$$M := \min\{k : Y_k - Z_k < 1\},$$
$$N := \min\{k : Y_k < 1\}.$$

Then $L_k = Y_k - Z_k$ and $\tilde{H}_x = M$, while $N = H_x$, the height of the corresponding one-sided interval tree under the simpler rule of Theorem 2.1. We prove some lemmas.

Lemma 3.1. *There exists a constant $A < \infty$, depending on the distribution of R but not on z, such that for every $z > 0$,*

$$\mathbb{E}\left(\sum_{j \geq 0:\, e^{S_j} \leq z} e^{S_j}\right) \leq Az. \tag{3.3}$$

Proof. Let $t = \ln z$. If U is the renewal function of $\{S_n\}_0^\infty$, i.e. the sum of the distribution functions of S_n, $n \geq 0$, then

$$z^{-1} \mathbb{E} \sum_{j:\, e^{S_j} \leq z} e^{S_j} = \mathbb{E}\left(\sum_{j \geq 0} e^{S_j - t}\mathbf{1}[S_j \leq t]\right) = \int_0^t e^{-(t-s)}\, dU(s).$$

By the Key Renewal Theorem [1, Theorem II.4.3], this converges, as $t \to \infty$, to $\mu^{-1}\int_0^\infty e^{-s}\, ds = \mu^{-1} < \infty$. Hence (3.3) holds for large z, say $z \geq z_0$, with $A = 1 + \mu^{-1}$. Evidently, (3.3) holds for $1 \leq z \leq z_0$ too, if A is large enough, and the case $z < 1$ is trivial. $\qquad\square$

Lemma 3.2. *There exists a constant $B < \infty$, depending on the distribution of R but not on y, such that for every $y > 0$,*

$$\mathbb{P}(Y_M \geq y) \leq B/y. \tag{3.4}$$

Proof. Let $y > 1$ and assume that $y \leq Y_M \leq 2y$. Since $L_M = Y_M - Z_M < 1$, then $Z_M > Y_M - 1 \geq y - 1$. Moreover, for $j \leq M$,

$$e^{S_j} \leq e^{S_M} = \frac{x}{Y_M} \leq \frac{x}{y}.$$

Hence, with $z = x/y$,

$$y - 1 \leq Z_M = \frac{Y_M}{x}\sum_{j=0}^{M-1} e^{S_j} \leq \frac{2y}{x}\sum_{j:\, e^{S_j} \leq z} e^{S_j}.$$

Consequently, taking the expectation and using Lemma 3.1,

$$(y-1)\,\mathbb{P}(y \leq Y_M \leq 2y) \leq \frac{2y}{x}\mathbb{E}\sum_{j:\, e^{S_j} \leq z} e^{S_j} \leq \frac{2y}{x}Az = 2A.$$

For $y \geq 2$, this yields

$$\mathbb{P}(y \leq Y_M \leq 2y) \leq \frac{2A}{y-1} \leq \frac{4A}{y},$$

and hence

$$\mathbb{P}(Y_M \geq y) = \sum_{j=0}^{\infty} \mathbb{P}\big(2^j y \leq Y_M < 2^{j+1} y\big) \leq \sum_{j=0}^{\infty} \frac{4A}{2^j y} = \frac{8A}{y},$$

which shows (3.4) if $B \geq 8A$. For $y < 2$, (3.4) is trivial provided $B \geq 2$. $\qquad\square$

Lemma 3.3. *There exists a constant $C < \infty$, depending on the distribution of R but not on x, such that for every $x > 0$,*

$$\mathbb{E}(H_x - \tilde{H}_x) \leq C,$$
$$\mathbb{E}(H_x - \tilde{H}_x)^2 \leq C.$$

Proof. Observe that if we continue after M, then the lenghts Y_{M+i}, $i = 0, \ldots, N - M$, are the interval lengths for a one-sided interval tree as in Theorem 2.1 with an initial (random) length Y_M. Hence, conditioned on $Y_M = y$, $N - M$ has the distribution of H_y in Theorem 2.1. Since (2.2) and (2.3) imply, for some $C' < \infty$ and all $y \geq 0$,

$$\mathbb{E}\, H_y \leq C'(1 + \ln_+ y),$$
$$\mathbb{E}(H_y)^2 \leq C'(1 + \ln_+ y)^2,$$

it follows that

$$\mathbb{E}(H_x - \tilde{H}_x) = \mathbb{E}(N - M) \leq C'\, \mathbb{E}(1 + \ln_+ Y_M),$$
$$\mathbb{E}(H_x - \tilde{H}_x)^2 = \mathbb{E}(N - M)^2 \leq C'\, \mathbb{E}(1 + \ln_+ Y_M)^2.$$

By Lemma 3.2, these expectations are bounded by some constant C. $\qquad\square$

Using Lemma 3.3, Theorem 2.1 easily implies Theorem 2.2: (2.5) follows from (2.2), while (2.6) follows from (2.3) and Minkowski's inequality in the form $|(\operatorname{Var} H_x)^{1/2} - (\operatorname{Var} \tilde{H}_x)^{1/2}| \leq \big(\mathbb{E}(H_x - \tilde{H}_x)^2\big)^{1/2}$; finally, Lemma 3.3 implies $(H_x - \tilde{H}_x)/\ln x \xrightarrow{\mathrm{p}} 0$, and Cramér's theorem yields (2.7) from (2.4) $\qquad\square$

4 Examples

We begin by considering the five versions studied by Itoh and Mahmoud [2], giving alternative proofs of their results. (Recall that they state the results in terms of $T_x = 2H_x + 1$, their S_x.)

We let $h_n := \sum_{k=1}^{n} 1/k$ denote the harmonic numbers. Similarly, $h_n^{(2)} := \sum_{k=1}^{n} 1/k^2$.

Example 4.1 (left preference). If we each time choose the left subinterval, $R = U$. Hence $X = -\ln(U)$, which has an exponential distribution $\text{Exp}(1)$. In particular, $\mu = \mathbb{E}\,X = 1$ and $\sigma^2 = \text{Var}\,X = 1$, and thus Theorem 2.1 yields, as shown in [2],

$$\mathbb{E}\,H_x \sim \ln x,$$
$$\text{Var}\,H_x \sim \ln x,$$

and the asymptotic normality

$$\frac{H_x - \ln x}{\sqrt{\ln x}} \xrightarrow{\text{d}} N(0, 1).$$

In this case, we can say more. Since $X_1, X_2, \ldots$ are i.i.d. random variables with an $\text{Exp}(1)$ distribution, their partial sums $S_1, S_2, \ldots$ are the points of a Poisson process with intensity 1 on $(0, \infty)$. Since H_x equals 1 plus the number of these points in $(0, \ln x]$, we see that, for every $x \geq 1$, $H_x - 1$ has a Poisson distribution

$$H_x - 1 \in \text{Po}(\ln x), \qquad x \geq 1.$$

This is also implicit in [2]; it is equivalent to the formula

$$\mathbb{E}\,e^{tT_x} = \mathbb{E}\,e^{t(2H_x+1)} = e^{3t}x^{e^{2t}-1}$$

derived there.

In particular, we have, for $x \geq 1$, the exact formulas $\mathbb{E}\,H_x = \ln x + 1$ and $\text{Var}\,H_x = \ln x$, or equivalently $\mathbb{E}\,T_x = 2\ln x + 3$ and $\text{Var}\,T_x = 4\ln x$, as shown in [2].

Note also that the logarithms of the lengths of the intervals corresponding to internal nodes are $\{\ln x - S_n\}_{n=0}^{H_x-1}$; omitting the point $\ln x$, these random points form a Poisson process in $[0, \ln x]$ with intensity 1. By a change of variables, the lengths of the intervals corresponding to internal nodes, except the root (which has length x), form a Poisson process in $[1, x]$ with intensity dy/y. Equivalently, the

division points except the last (which is less than 1) form a Poisson process in $[1, x]$ with intensity dy/y.

Let us end this example by noting that the left preference one-sided interval tree is related to records in an i.i.d. sequence, as studied by Rényi [5] and many others. Indeed, let x be an integer and let $\xi_1, \ldots, \xi_x$ be a sequence of independent random variables with a common continuous distribution. (As is well-known, it is equivalent to consider a a random permutation of $\{1, \ldots, x\}$.) The set of record times is $\{i : \xi_i > \xi_j \text{ for } j < i\}$. To find the records, we may start by finding the largest value ξ_i, noting that its index i is uniform over $\{1, \ldots, x\}$. This index i is the last record time, so to find the other records, we continue recursively in the interval $\{1, \ldots, i-1\}$ to the left of it. We thus find the largest value there, continue with the interval to the left of it, and so on, until the left interval is empty. Consequently, the records are found by a discrete version of the left preference one-sided interval tree (or perhaps rather of the car-parking version of it), with the height corresponding to the number of records.

It is therefore not surprising that the results above for the left preferecne one-sided interval tree have analogues for records; for comparison we briefly quote some well-known results, see e.g. [5], [6] for details and proofs. For example, the expected number of records in $\{1, \ldots, x\}$ is the harmonic number $h_x = \ln x + O(1)$, the variance is $h_x - h_x^{(2)} = \ln x + O(1)$, the distribution is asymptotically normal and is well approximated by a $\mathrm{Po}(\ln x)$ distribution. More precisely, if we let $I_i = 1$ when i is a record time and $I_i = 0$ otherwise, then $\mathbb{P}(I_i = 1) = 1/i$, and the random variables I_i are independent. The number of records equals $\sum_{i=1}^{x} I_i$, which yields the results just stated. It follows also that the random set of record times can be obtained from the points of a Poisson process on $(0, x)$ with intensity dy/y by rounding each point to the nearest larger integer (ignoring repetitions). After suitable rescaling (division by x), the set of record times thus converges to a Poisson process on $(0, 1)$ with intensity dy/y.

Example 4.2 (min preference). If we each time choose the smallest interval, $R = \min(U, 1 - U)$. Thus R has a uniform distribution on $[0, 1/2]$, and $2R \stackrel{\mathrm{d}}{=} U$. Hence $X = -\ln(R) = Y + \ln 2$ with $Y = -\ln(2R) \in \mathrm{Exp}(1)$.

We have $\mu = \mathbb{E} X = 1 + \ln 2$ and $\sigma^2 = \mathrm{Var} X = \mathrm{Var} Y = 1$.

Consequently, Theorem 2.1 shows that H_x is asymptotic normal with

$$\mathbb{E}\, H_x = \frac{1}{1 + \ln 2}\ln x + \frac{1}{2(1 + \ln 2)^2} + \frac{1}{2} + o(1),$$

$$\mathrm{Var}\, H_x \sim \frac{1}{(1 + \ln 2)^3}\ln x.$$

We have recovered the result of [2], and found a sharper estimate of $\mathbb{E}\, H_x$.

Example 4.3 (max preference). If we each time choose the largest interval, $R = \max(U, 1 - U)$. Thus R has a uniform distribution on $[1/2, 1]$. Hence $X = -\ln(R)$ has the density $2e^{-x}$ on $[0, \ln 2]$. (Here X is an $\mathrm{Exp}(1)$ variable conditioned to be less that $\ln 2$.) A simple calculation yields $\mu = 1 - \ln 2$ and $\sigma^2 = 1 - 2\ln^2 2$. Consequently, H_x is asymptotic normal with

$$\mathbb{E}\, H_x = \frac{1}{1 - \ln 2}\ln x + \frac{2 - 2\ln 2 - \ln^2 2}{2(1 - \ln 2)^2} + o(1),$$

$$\mathrm{Var}\, H_x \sim \frac{1 - 2\ln^2 2}{(1 - \ln 2)^3}\ln x.$$

We have again recovered the result of [2] with a sharper estimate of $\mathbb{E}\, H_x$.

Example 4.4 (proportionate preference). If we choose between the two subintervals with probabilities proportional to their lengths, we have $p(u) = u$ and (2.1) shows that R has the density function $2u$, $0 < u < 1$. Thus, $\mathbb{P}(R \le u) = u^2$, $0 \le u \le 1$, and for $y \ge 0$,

$$\mathbb{P}(X > y) = \mathbb{P}\big(R < \exp(-y)\big) = \exp(-2y).$$

In other words, $X = -\ln(R)$ has an exponential distribution $\mathrm{Exp}(1/2)$ with mean $\mu = 1/2$ and variance $\sigma^2 = 1/4$.

Since X has an exponential distribution, we have as in Example 4.1 that the partial sums $S_1, S_2, \ldots$ are the points of a Poisson process $(0, \infty)$, this time with intensity 2. Hence, for every $x \ge 1$, $H_x - 1$ has a Poisson distribution

$$H_x - 1 \in \mathrm{Po}(2\ln x), \qquad x \ge 1.$$

(This is implicit in [2].) In particular, we have, for $x \ge 1$, the exact formulas $\mathbb{E}\, H_x = 2\ln x + 1$ and $\mathrm{Var}\, H_x = 2\ln x$, or equivalently $\mathbb{E}\, T_x = 4\ln x + 3$ and $\mathrm{Var}\, T_x = 8\ln x$.

Similarly to Example 4.1, the lengths of the intervals corresponding to internal nodes, except the root (which has length x), form a Poisson process in $[1, x]$ with intensity $2 dy/y$,

Note that one way to generate a proportionate one-sided interval tree is to choose a single random point Y (with uniform distribution) in $(0, x)$, and then always choose the subinterval containing Y. Since the conditional distribution of Y given the first k steps in the building of the one-sided interval tree remains uniform on the next interval to be split, this gives the required independence between different steps, in spite of the fact that the choice of Y is made only once. Seen in this way, the proportionate one-sided interval tree becomes a continuous analogue of the FIND algorithm for finding the element of a given rank in a set of distinct elements, if we choose the rank at random. (Recall that FIND recursively chooses a random pivot, compares the other elements to the pivot, and continues with either the elements larger or the elements smaller than the pivot until the element of the right rank is found.)

Example 4.5 (uniform preference). If we each time choose one of the subintervals at random, we take R equal to either U or $1 - U$, with probability $1/2$. This choice is independent of U, and since U and $1 - U$ have the same distribution, it follows that $R \stackrel{\mathrm{d}}{=} U$. (This also follows from (2.1), with $p(u) = 1/2$.) Thus R is as in Example 4.1, and hence all results there hold for uniform preference too, as observed by Itoh and Mahmoud [2].

We add one more example of the same type.

Example 4.6 (anti-proportionate preference). If we do the opposite to Example 4.4 and choose the subinterval to be *discarded* with probabilities proportional to the lengths of the subinterval, we have $p(u) = 1 - u$ and (2.1) shows that R has the density function $2(1 - u)$, $0 < u < 1$. Simple calculations show that X has the density function $2e^{-x} - 2e^{-2x}$, with mean $\mu = 3/2$ and variance $\sigma^2 = 5/4$. Consequently, H_x is asymptotic normal with

$$\mathbb{E}\, H_x = \frac{2}{3} \ln x + \frac{7}{9} + o(1),$$

$$\mathrm{Var}\, H_x \sim \frac{10}{27} \ln x.$$

It follows from the results of [3], see Example 4.8 below, that actually, for every $x \geq 1$,

$$\mathbb{E}\, H_x = \frac{2}{3} \ln x + \frac{7}{9} + \frac{2}{9} x^{-3}. \tag{4.1}$$

Example 4.7 (car-parking). In all the examples with different policies above, the corresponding car-parking version has, by Theorem 2.2, the same asymptotic normal distribution and the same leading terms in the asymptotics of the mean and variance of $\tilde{H}_x$. This was shown in [2] for the mean of $\tilde{H}_x$ (there denoted S_x) for left (or uniform) preference.

Example 4.8 (one-sided m-ary interval trees). Javanian, Mahmoud and Vahidi-Asl [3] studied a generalization to one-sided m-ary interval trees, where intervals are split into m subintervals, using $m - 1$ independent and uniformly distributed division points. Here $m \geq 2$ is a fixed integer; $m = 2$ gives the binary splitting considered above. In the case studied in [3], one of the subintervals is chosen at random, according to some arbitrary distribution but independently of the interval lengths. As observed in [3], all such selection rules give the same result, since the m spacings defined by $m - 1$ independent uniformly distributed division points in an interval are exchangeable. We may thus just as well assume that we always select the left interval and continue by dividing it until we obtain an interval with length less than 1.

This can be regarded as another instance of Theorem 2.1, where R now is the smallest of $m-1$ independent random variables $U_1, \ldots, U_{m-1}$ with a uniform distribution on $(0, 1)$. As is well-known, for $0 \leq u \leq 1$,

$$\mathbb{P}(R > u) = \prod_{i=1}^{m-1} \mathbb{P}(U_i > u) = (1 - u)^{m-1},$$

and thus R has the density function $(m - 1)(1 - u)^{m-2}$, $0 \leq u \leq 1$. (A Beta$(1, m - 1)$ distribution.) To find the mean and variance of $X = -\ln R$, we compute its moment generating function: for $t < 1$,

$$\mathbb{E}\, e^{tX} = \mathbb{E}\, e^{-t \ln R} = \mathbb{E}\, R^{-t} = \int_0^1 u^{-t}(m - 1)(1 - u)^{m-2}\, du$$

$$= (m - 1)B(1 - t, m - 1) = (m - 1)\frac{\Gamma(1 - t)\Gamma(m - 1)}{\Gamma(m - t)}$$

$$= \frac{(m - 1)!}{(1 - t)\cdots(m - 1 - t)} = \prod_{i=1}^{m-1} \frac{1}{1 - t/i}.$$

Hence X has the same distribution as $\sum_{i=1}^{m-1} Y_i$, with $Y_i \in \text{Exp}(1/i)$

independent. Consequently (or by differentiating $\mathbb{E}\,e^{tX}$ at $t=0$),

$$\mu = \mathbb{E}\,X = \sum_{i=1}^{m-1}\mathbb{E}\,Y_i = h_{m-1},$$

$$\sigma^2 = \operatorname{Var} X = \sum_{i=1}^{m-1}\operatorname{Var} Y_i = h_{m-1}^{(2)}.$$

Consequently, by Theorem 2.1, as shown in [3] by another method, H_x is asymptotically normal with

$$\mathbb{E}\,H_x = \frac{1}{h_{m-1}}\ln x + \frac{h_{m-1}^{(2)}}{2h_{m-1}^2} + \frac{1}{2} + o(1),$$

$$\operatorname{Var} H_x \sim \frac{h_{m-1}^{(2)}}{h_{m-1}^3}\ln x.$$

Javanian, Mahmoud and Vahidi-Asl [3] further gave an exact formula for $\mathbb{E}\,H_x$.

For $m=2$, we have the left preference binary one-sided interval tree treated in Example 4.1. For $m=3$, the density of R is the same as in the anti-proportionate binary one-sided interval tree treated in Example 4.6. Hence the heights H_x have the same distribution in both cases. In particular, (4.1) follows from the exact formula for $\mathbb{E}\,H_x$ in [3].

Several variations are possible. For example, we can choose the selected subinterval according to some other policy, depending only on the relative lengths. We give the results for two cases, omitting the calculations.

For the proportionate version, with subintervals chosen with probabilities proportional to their lengths, it is easily seen that R has the density $m(m-1)u(1-u)^{m-2}$, $0 \le u \le 1$, and a calculation of the moment generating function $\mathbb{E}\,e^{tX} = \mathbb{E}\,R^{-t}$ similar to the calculation above shows that X now has the same distribution as $\sum_{i=2}^{m}Y_i$, with $Y_i \in \operatorname{Exp}(1/i)$ independent as above. Consequently, $\mu = h_m - 1$ and $\sigma^2 = h_m^{(2)} - 1$. Thus, by Theorem 2.1, for the preference m-ary one-sided interval tree generalizing Example 4.4, H_x is asymptotically normal with

$$\mathbb{E}\,H_x = \frac{1}{h_m - 1}\ln x + \frac{h_m^{(2)} - 1}{2(h_m - 1)^2} + \frac{1}{2} + o(1),$$

$$\operatorname{Var} H_x \sim \frac{h_m^{(2)} - 1}{(h_m - 1)^3}\ln x.$$

For the minimal version, R has the density $m(m-1)(1-mu)^{m-2}$ on $0 \leq u \leq 1/m$. Hence R has the same distribution as R'/m, where R' is the corresponding variable in the left preference case. Consequently, $X = X' + \ln m$, where $X' = \sum_{i=1}^{m-1} Y_i$ with Y_i as above, and we now have $\mu = h_{m-1} + \ln m$ and $\sigma^2 = h_{m-1}^{(2)}$. Thus, by Theorem 2.1, for the min preference m-ary one-sided interval tree generalizing Example 4.2, H_x is asymptotically normal with

$$\mathbb{E}\, H_x = \frac{1}{h_{m-1} + \ln m} \ln x + \frac{h_{m-1}^{(2)}}{2(h_{m-1} + \ln m)^2} + \frac{1}{2} + o(1),$$

$$\mathrm{Var}\, H_x \sim \frac{h_{m-1}^{(2)}}{(h_{m-1} + \ln m)^3} \ln x.$$

Another variation, covered by Theorem 2.2, is to consider parking where $m-1$ cars, each of length $1/(m-1)$, park at the same time.

Acknowledgement

References

[1] Gut, A. (1988), Stopped Random Walks. New York: Springer-Verlag.

[2] Itoh, Y. and Mahmoud, H. (2003), One-sided variations on interval trees. J. Appl. Prob., **40**, No. 3, 654–670.

[3] Javanian, M., Mahmoud, H., and Vahidi-Asl, M. (2004), Paths in m-ary interval trees. Discrete Mathematics, **287**, 45-53.

[4] Rényi, A. (1958), On a one-dimensional random space-filling problem. (Hungarian.) Magyar Tud. Akad. Mat. Kutató Int. Közl. **3**, 109–127. English transl. in Selected Papers of Alfréd Rényi, Vol. II: 1956–1961. Ed. Pál Turán. Akadémiai Kiadó, Budapest, 1976, 173–188.

[5] Rényi, A. (1962), On the extreme elements of observations. (Hungarian) Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. **12**, 105–121.

English transl. in Selected Papers of Alfréd Rényi, Vol. III: 1962–1970. Ed. Pál Turán. Akadémiai Kiadó, Budapest, 1976, 50–66.

[6] S.I. Resnick, S. I. (1987), Extreme Values, Regular Variation and Point Processes. New York: Springer-Verlag.