Acta Informatica 26, 123–130 (1988)

Applications of the Theory of Records in the Study of Random Trees *

Luc Devroye

School of Computer Science, McGill University, 805 Sherbrooke Street West, Montreal, Canada H3A 2K6

Summary. The theory of records in sequences of independent identically distributed random variables leads to simple proofs of various properties of random trees, including among other things, the limit law for the depth of the last node of random ordered trees, random union-find trees, and random binary search trees.

1. Introduction

In this note, we point out the connection between random trees and records in a sequence of independent identically distributed (iid) random variables. This allows us to obtain short and hopefully insightful proofs of a number of properties of these trees.

To illustrate this, we will prove that the length of the path from the root to the last node added in a tree with n elements is asymptotically normally distributed with mean and variance both equal to $2 \log n$. This property was obtained independently by Mahmoud and Pittel (see announcement of the result in Mahmoud and Pittel [15, footnote 1]) based on a close analysis of the behavior of Stirling numbers of the first kind. We also consider models for random ordered trees and random union-find trees.

2. Records. The Connection

In an iid sequence $X_1, \ldots, X_n$ of random variables with a given density, we say that X_i is a *record* (or up-record) if $X_i = \max(X_1, \ldots, X_i)$, and we let Y_i be the indicator of this event. Records have a few useful properties that are perhaps best captured in the survey article of Glick [9]. Notice first that the distribution of $(Y_1, \ldots, Y_n)$ (and of all the functions of these random variables) does not depend upon the common density of the X_i's. Also, we may without

* Research of the author was sponsored by NSERC Grant A 3456 and by FCAC Grant EQ-1678

loss of generality consider that $X_1, \ldots, X_n$ is a random equiprobable permutation of $1, \ldots, n$. The basic properties of records can be summarized as follows:

A) If $R_1, \ldots, R_n$ are the partial ranks of $X_1, \ldots, X_n$ (i.e. $R_i=j$ if and only if X_i is the j-th largest among $X_1, \ldots, X_i$), then $R_1, \ldots, R_n$ are independent. This implies that $Y_1, \ldots, Y_n$ are independent. Also, for each j, R_j is uniformly distributed on $1, \ldots, j$.

B) $P(Y_i=1)=1/i$, all $i \geqq 1$.

C) The position (or index) of the last record in $X_1, \ldots, X_n$ is uniformly distributed on $1, \ldots, n$.

D) If N_n is the number of records in $X_1, \ldots, X_n$, H_n is the summed harmonic series $\sum_{i=1}^{n} 1/i$, and $H_n^{(2)}$ is $\sum_{i=1}^{n} 1/i^2$, then $E(N_n)=H_n$ and $Var(N_n)=H_n-H_n^{(2)}$.

E) $N_n/\log n \to 1$ in probability.

F) $(N_n-\log n)/\sqrt{\log n}$ tends in distribution to a normal random variable as $n\to\infty$ (Renyi [17]).

Remark 1. About the proofs of D, E, and F. Property D is immediate from A and B. Using estimates for the harmonic series found e.g. in Knuth [11, 12], we note in particular that $E(N_n)-\log n \to \gamma = 0.57721566499\ldots$ (γ is Euler's constant), and $Var(N_n)-\log n \to \gamma-\pi^2/6$. The law of large numbers (E) follows from D and Chebyshev's inequality. The central limit theorem F can be obtained very easily by verifying the conditions of the Lindeberg-Feller central limit theorem (see e.g. Chow and Teicher [5, p. 291]). The zero mean random variables Y_i-1/i satisfy the central limit theorem $\left(\text{i.e. } \sum_{i=1}^{n}(Y_i-1/i)/s_n\right.$ tends in distribution to the normal law as $n\to\infty$ where $\left. s_n^2 \stackrel{\Delta}{=} \sum_{i=1}^{n} Var(Y_i)\right)$ when

$$\sum_{j=1}^{n} \int_{|x|>\varepsilon s_j} x^2 \, \mathrm{d}F_j(x) = o(s_n^2), \qquad \text{all } \varepsilon>0,$$

(the Lindeberg-Feller condition) where F_j is the distribution function of Y_j-1/j. In our case, we have $s_n^2=\sum_{i=1}^{n}\frac{1}{i}\left(1-\frac{1}{i}\right)=H_n-H_n^{(2)}$. Note that the Lindeberg-Feller condition is satisfied since $s_n\to\infty$, and each F_j puts all its mass on $[-1, 1]$. Thus, we conclude that $(N_n-H_n)/\sqrt{H_n-H_n^{(2)}}$ tends in distribution to a normal random variable. □

Remark 2. Generating functions. Records can also be handled via characteristic functions or generating functions. From the independence of the Y_i's, we see that the characteristic function $\phi_n(t)$ of N_n is

$$\phi_n(t)=E\left(e^{it\sum_{j=1}^{n} Y_j}\right)=\prod_{j=1}^{n} E(e^{itY_j})=\prod_{j=1}^{n}\left(\frac{e^{it}}{j}+1-\frac{1}{j}\right).$$

Similarly, the generating function $f_n(z)$ $\left(\text{defined as } \sum_{j=0}^{\infty} P(N_n=j)\,z^j\right)$ can be obtained by replacing e^{it} by z in the expression for $\phi_n(t)$. The standard combinatorial way of deriving $f_n(z)$ uses the recursion

$$P(N_n=i+1)=\frac{1}{n}\,P(N_{n-1}=i)+\frac{n-1}{n}\,P(N_{n-1}=i+1)$$

subjected to the boundary value $P(N_1=1)=1$. Hofri [10, pp. 112–117] and Sedgewick [10, pp. 151–152] used this technique to study the number of records in a random permutation of 1, ..., n. The coefficient z^i in the expansion of $f_n(z)$ yields

$$P(N_n=i)=\frac{1}{n!}\begin{bmatrix} n \\ i \end{bmatrix},$$

where $[.]$ denotes the Stirling number of the first kind. This provides the well-known interpretation of the Stirling number $\begin{bmatrix} n \\ i \end{bmatrix}$ as the number of permutations of n numbers having precisely i records. The moments of N_n can be determined very quickly from the characteristic or generating functions. We obtain $E(N_n)=H_n$ and $Var(N_n)=H_n-H_n^{(2)}$, confirming item D above. We will not use the generating functions in any of our derivations, basing all our results solely on properties A – F stated above. □

3. Random Ordered Trees

An ordered tree on n nodes is a tree in which each node can have any number of children, and the children are ordered from oldest to youngest. A *random ordered* tree can be constructed incrementally by starting with a root node, and given a tree with i nodes, attaching the $i+1$-st node as the youngest child of a random equiprobable node among the i nodes already present. Note for example that the number of children of node i is distributed as

$$\sum_{j=i+1}^{n} I_{[\text{node } j \text{ attaches itself to node } i]}.$$

It is obvious that the expected value is $\sum_{j=i+1}^{n} \frac{1}{j-1}=H_{n-1}-H_{i-1}$. In particular, the number of offspring is distributed as the number of records in an iid sequence $X_1, \ldots, X_n$, with index greater than i. For fixed i, the number of children is thus asymptotically distributed like $\log n+N\sqrt{\log n}$ where N is a normal random variable.

Consider next the *level* L_i of node i in the random ordered tree, i.e. the distance from the root to node i in the tree. Clearly, $L_1=0$. Also, L_i is stochastically smaller than L_j for $i<j$. Many of the structural properties of the tree are related to the behavior of the random variables L_i, $1\leq i\leq n$.

Theorem O1. *L_n is distributed as the number of records in a random equiprobable permutation of* $1, \ldots, n-1$. *Hence,*

A) $E(L_n) = H_{n-1}$.

B) $Var(L_n) = H_{n-1} - H_{n-1}^{(2)}$.

C) $L_n/E(L_n) \to 1$ *in probability as* $n \to \infty$.

D) $(L_n - E(L_n))/\sqrt{Var(L_n)}$ *converges in distribution to a normal random variable.*

Proof of Theorem O1. It suffices to prove the first statement of the Lemma, and to apply the properties of records stated in Sect. 2. The father node of element n is a uniform number on $1, \ldots, n-1$, say a_1. Given a_1, its father node a_2, if $a_1 > 1$, is in turn uniformly distributed on $1, \ldots, a_1 - 1$. We continue this until we encounter $a_m = 1$, the root of the tree. The level of node n is then equal to m. Consider next a totally different experiment based upon $n-1$ iid uniform [0, 1] random variables $X_1, \ldots, X_{n-1}$. Let b_1 be the index of the last record. We know from property C that b_1 is uniformly distributed on $1, \ldots, n-1$. Given b_1, let b_2 be the index of the last record (the maximum) of $X_1, \ldots, X_{b_1}$. Clearly, given $b_1 > 1$, b_2 is uniformly distributed on $1, \ldots, b_1 - 1$. We keep going until we find for the first time $b_m = 1$. Clearly, the number of records in the sequence is m. By comparison of the two experiments, we can now conclude that L_n, the level of node n in the random ordered tree, is distributed as the number of records in $X_1, \ldots, X_{n-1}$. □

4. Random Union-Find Trees

Consider n singleton sets, each consisting of a different element. Grab two sets uniformly and at random, and join them using the usual union-find tree structure (see e.g. Aho, Hopcroft, and Ullman [2]). Repeat this operation until all sets are joined into one set of n elements. The corresponding tree is called the union-find tree; we assume that in the join operation, each set is equally likely to end up in the subtree of the root of the other set. This model (called the random sets model by Sedgewick [19]) is due to Doyle and Rivest [8]. For other models, see e.g. Yao [20] and Knuth and Schonhage [13].

Let $L_1, \ldots, L_n$ be the levels of the elements in the union-find tree. By symmetry, we see that the L_i's are identically distributed. The level of element one is initially zero since it starts as a root of a singleton set. Every time the set to which this element belongs is selected as the set to be joined as a subtree of the root of another set, its level increases by one. Call this event E_i if it happens during the iteration in which there are i sets left. Thus, each L_i is distributed as

$$\sum_{i=2}^{n} I_{E_i},$$

where the random variables I_{E_i} are independent, and Bernoulli with success probability $1/i$. In the notation of random permutations (see Sect. 2), this is distributed as $\sum_{i=2}^{n} I_{R_i=i}$. We conclude that L_1+1 is distributed as N_n, the number of records in an iid sequence of length n. Hence, properties D, E and F of records apply to each L_i+1. See also Theorem 8 of Devroye [7].

5. Random Binary Search Trees

Consider a random binary search tree on n nodes constructed in the usual manner from a random equiprobable permutation of $1, \ldots, n$ (see Aho, Hopcroft and Ullman [1, 2] for definitions and properties). These trees were studied by Lynch [14], Knuth [11], Robson [18], Sedgewick [19], Pittel [16], Mahmoud and Pittel [15], Brown and Shubert [3], Devroye [6, 7] and others.

We will show

Theorem S1. *Let L_n be the level of the last node added to a random binary search tree. Then*

A) $E(L_n)=2H_n-2$.

B) $Var(L_n)=2H_n-4H_n^{(2)}+2$.

C) $L_n/E(L_n)\to 1$ *in probability as* $n\to\infty$.

D) $(L_n-E(L_n))/\sqrt{Var(L_n)}$ *converges in distribution to a normal random variable.*

Proof of Theorem S1. We consider a random binary search tree constructed from a random equiprobable permutation $X_1, \ldots, X_n$ of $1, \ldots, n$. Let L and R partition the X_i sequence $(1\leqq i<n)$ into two sequences $X_{l1}, \ldots, X_{lk}$ and $X_{r1}, \ldots, X_{rm}$, where $k+m=n-1$, and each X_{li} is smaller than X_n, and each X_{ri} is larger than X_n. In fact, $k=X_n-1$ and $m=n-X_n$. Within L and R, the order of appearances of the X_i's is unaltered, however. The crucial observation we make now is that the level L_n of X_n is the sum of the number of records (up-records) in L plus the number of down-records in R. Equally crucial is that, given k, the L sequence is a random equiprobable permutation of $1, \ldots, k$, and similarly for the R sequence. Also, the two random permutations are independent. Thus, the number L_n is distributed, by mirroring, as the sum of the number of down-records in L plus the number of up-records in R. Consider now the sequence $X_n, X_1, X_2, \ldots, X_{n-1}$, with partial ranks $R_1, \ldots, R_n$. Notice the slight off-set of the indices: R_1 is the partial rank of X_n, R_2 is the partial rank of X_1, and so forth. We conclude that L_n is distributed as

$$\sum_{i=2}^{n} I_{R_i\in\{1,i\}}.$$

By the independence of the R_i's, we can now easily deduce all the stated properties. In particular,

$$E(L_n)=\sum_{i=2}^{n}\frac{2}{i}=2H_n-2.$$

Next,

$$Var(L_n)=\sum_{i=2}^{n}\frac{2}{i}\left(1-\frac{2}{i}\right)=(2H_n-2)-(4H_n^{(2)}-4)=2H_n-4H_n^{(2)}+2.$$

By Chebyshev's inequality, $L_n/E(L_n)\to 1$ in probability, and by the Lindeberg-Feller central limit theorem (see Remark 1), $(L_n-E(L_n))/\sqrt{Var(L_n)}$ converges in distribution to a normal random variable. □

Theorem S1 does not contain elements that were not known before. The novelty is in the proof: the generating function approach was used by many before: the generating function of L_n is $\prod_{j=2}^{n}\frac{j-2+2z}{j}$ (see e.g. Sedgewick [19], p. 143). By independence of the R_i's, observe that this is indeed the generating function of $\sum_{i=2}^{n} I_{R_i\in\{1,i\}}$. Lynch [14] and Knuth [11] proved that

$$P(L_n=k)=\frac{1}{n!}\begin{bmatrix} n-1 \\ k \end{bmatrix} 2^k, \qquad 1\leqq k\leqq n.$$

The generating function readily yields the exact values of $E(L_n)$ and $Var(L_n)$ given in Theorem S1 (Sedgewick [19], p. 144). Chebyshev's inequality allows one to conclude the weak law of large numbers given in Theorem S1. See also Mahmoud and Pittel [15]. The limit law (part D of Theorem S 1) is announced in footnote 1 of Mahmoud and Pittel [15], and is obtained by these authors based on a finer analysis of the Stirling numbers of the first kind.

6. Large Deviation Results

The random variable L_n studied in this note was in each case related to a sum of independent random variables related to record times in an i.i.d. sequence. Hence, it is straightforward to exploit this fact to obtain large deviation inequalities. For example, if N_n is the number of records in a random permutation of $1, \ldots, n$, then we have:

Theorem L1. *For every* $\varepsilon>0$,

$$P(N_n-H_n\geqq\varepsilon)\leqq e^{\varepsilon-(H_n+\varepsilon)\log\left(1+\frac{\varepsilon}{H_n}\right)}\leqq e^{-\frac{\varepsilon^2}{2H_n+\varepsilon}}$$

and

$$P(N_n-H_n\leqq-\varepsilon)\leqq e^{-\varepsilon-(H_n-\varepsilon)\log\left(1-\frac{\varepsilon}{H_n}\right)}\leqq e^{-\frac{\varepsilon^2}{2H_n}}.$$

Proof of Theorem L1. Let $t>0$ be arbitrary. We use Chernoff's exponential bounding technique (Chernoff [4]) as follows:

$$P\left(\sum_{i=1}^{n}\left(Y_i-\frac{1}{i}\right)\geqq\varepsilon\right)\leqq e^{t\sum_{i=1}^{n}Y_i-t(H_n+\varepsilon)}=\prod_{i=1}^{n}\left(1-\frac{1}{i}+\frac{1}{i}e^t\right)e^{-t(H_n+\varepsilon)}$$

$$\leqq e^{H_n(e^t-1-t)-t\varepsilon}=e^{\varepsilon-(H_n+\varepsilon)\log\left(1+\frac{\varepsilon}{H_n}\right)}$$

when we take $t = \log(1+\varepsilon/H_n)$ (this choice minimizes the bound). This can be bounded from above by using the inequality $\log(1+u) \geqq 2u/(2+u)$, valid for all $u \geqq 0$. Similarly,

$$P\left(\sum_{i=1}^{n}\left(Y_i - \frac{1}{i}\right) \leqq -\varepsilon\right) \leqq e^{-t\sum_{i=1}^{n} Y_i + t(H_n - \varepsilon)} = \prod_{i=1}^{n}\left(1 - \frac{1}{i} + \frac{1}{i}e^{-t}\right) e^{t(H_n-\varepsilon)}$$

$$\leqq e^{H_n(t-1-e^{-t}) - t\varepsilon} = e^{-\varepsilon - (H_n - \varepsilon)\log\left(1 - \frac{\varepsilon}{H_n}\right)}$$

when we take $t = -\log(1-\varepsilon/H_n)$ (this choice minimizes the bound). This can be bounded from above by using the inequality $-(1-u)\log(1-u) \leqq u - u^2/2$, valid for $1 > u \geqq 0$. □

We have seen that for the random ordered tree, L_n is distributed as N_{n-1}, and that for the random union-find tree under the random sets model, all L_i's are distributed as $N_n - 1$. In both cases, Theorem L1 applies with the appropriate changes. For the random binary search tree, some changes are necessary, since $L_n - (2H_n - 2)$ is distributed as $\sum_{i=2}^{n}\left(Y_i - \frac{2}{i}\right)$, where the Y_i's are independent and Bernoulli with success probability $2/i$. The proof of Theorem L1 can now be mimicked.

References

1. Aho, A.V., Hopcroft, J.E., Ullman, J.D.: The design and analysis of computer algorithms. Reading, Mass.: Addison-Wesley 1975
2. Aho, A.V., Hopcroft, J.E., Ullman, J.D.: Data structures and algorithms. Reading, Mass.: Addison-Wesley 1983
3. Brown, G.G., Shubert, B.O.: On random binary trees. Math. Oper. Res. **9**, 43–65 (1984)
4. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist. **23**, 493–507 (1952)
5. Chow, Y.S., Teicher, H.: Probability theory. Berlin Heidelberg New York: Springer 1978
6. Devroye, L.: A note on the height of binary search trees. J. ACM **33**, 489–498 (1986)
7. Devroye, L.: Branching processes in the analysis of the heights of trees. Acta Informatica **24**, 277–298 (1987)
8. Doyle, J., Rivest, R.L.: Linear expected time of a simple union-find algorithm. Inform. Proc. Lett. **5**, 146–148 (1976)
9. Glick, N.: Breaking records and breaking boards. Am. Math. Monthly **85**, 2–26 (1978)
10. Hofri, M.: Probabilistic analysis of algorithms. Berlin Heidelberg New York: Springer 1987
11. Knuth, D.E.: The art of computer programming. Vol. 3: Sorting and searching. Reading, Mass.: Addison-Wesley 1973
12. Knuth, D.E.: The art of computer programming. Vol. 1: Fundamental algorithms. 2nd Edn. Reading, Mass.: Addison-Wesley 1973
13. Knuth, D.E., Schonhage, A.: The expected linearity of a simple equivalence algorithm. Theor. Comput. Sci. **6**, 281–315 (1978)
14. Lynch, W.C.: More combinatorial problems on certain trees. Comput. J. **7**, 299–302 (1965)
15. Mahmoud, H., Pittel, B.: On the most probable shape of a search tree grown from a random permutation. SIAM J. Algebr. Discr. Methods **5**, 69–81 (1984)
16. Pittel, B.: On growing random binary trees. J. Math. Anal. Appl. **103**, 461–480 (1984)

17. Renyi, A.: Theorie des elements saillant d'une suite d'observations. In: Colloquium on combinatorial methods in probability theory, pp. 104–115. Aarhus Universitet, Denmark: Mathematisk Institut 1962
18. Robson, J.M.: The height of binary search trees. Austr. Comput. J. **11**, 151–153 (1979)
19. Sedgewick, R.: Mathematical analysis of combinatorial algorithms. In: Louchard, G., Latouche, G. (eds.) Probability Theory and Computer Science, pp. 123–205. London: Academic Press 1983
20. Yao, A.C.: On the average behavior of set merging algorithms (extended abstract). In: Proceedings of the Eighth Annual Symposium on Theory of Computing, pp. 192–195 (1976)

Received June 25, 1984/March 25, 1988