BIT 20 (1980), 157–163

A NOTE ON THE STACK SIZE OF REGULARLY DISTRIBUTED BINARY TREES

R. KEMP

Abstract.

Assume that in one unit of time a node is stored in the stack or is removed from the top of the stack during postorder-traversing of a binary tree. If all binary trees are equally likely the average stack size after t units of time and the variance is computed as a function of the proportion $\varrho = t/n$.

1. Introduction.

Let $T(n)$, $n \in \mathrm{N}$, be the set of all *extended binary trees* ([6]) with n leaves and $T \in T(n)$. The *stack size* $S(T)$ is recursively defined by

$$S(T) := \text{IF } |T| = 1 \text{ THEN } 1 \text{ ELSE IF } S(T_1) > S(T_2) \text{ THEN } S(T_1) \text{ ELSE } S(T_2)+1;$$

where $|T|$ is the number of nodes of the tree T and T_1 (T_2) is the left (right) subtree of T. $S(T)$ is the maximum number of nodes stored in the stack during postorder-traversing of $T \in T(n)$. In [4] it is implicitly shown that the average stack size of a binary tree $T \in T(n)$ is asymptotically given by $\sqrt{(\pi n)} - 1/2 + O(\ln n/\sqrt{n})$ assuming that all n-node trees are equally likely. The variance is computed in [5] and is asymptotically given by

$$(\pi/3-1)\pi n + \tfrac{1}{12} - \tfrac{1}{18}\pi^2 + \tfrac{1}{12}\pi + O(\ln n/n^{\frac{1}{2}-\varepsilon}) \text{ for all } \varepsilon > 0 .$$

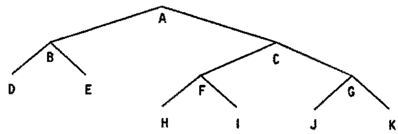
In this paper we consider an analogous problem. Evaluating a binary tree $T \in T(n)$ in postorder we assume that in one unit of time a node is stored in the stack or is removed from the top of the stack. Considering all trees $T \in T(n)$ equally likely we shall compute the average number of nodes $R_1(n,t)$ stored in the stack after t units of time as a function of the proportion $\varrho = t/n$. Moreover, we give an asymptotic equivalent for the sth moment $R_s(n,t)$ with respect to the origin, and for the variance.

2. The average stack size after t units of time.

Obviously, each path from $(t,k) = (1,1)$ to $(t,k) = (2n-1,1)$ in Figure 1 corresponds to the evaluation of a binary tree $T \in T(n)$ in postorder ([6; p. 316]); for example, the marked path in Figure 1 corresponds to the following tree $T \in T(6)$.

Part of this paper was presented at the 6th Colloquium on Automata, Languages and Programming, ICALP' 79, July 1979.

Received October 5, 1979.



If we reach the point (i,j) then we have exactly j nodes in the stack after i units of time.

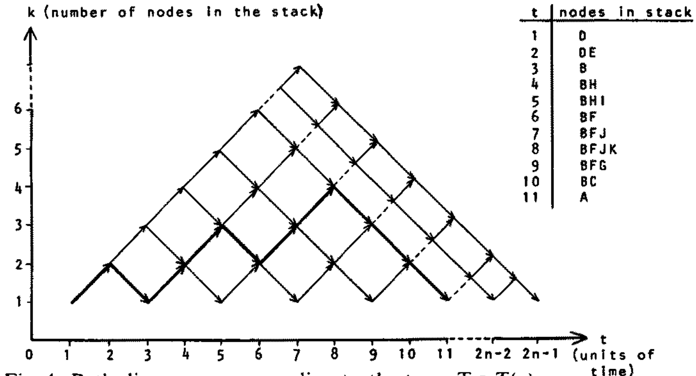


Fig. 1. Path diagram corresponding to the trees $T \in T(n)$.

Now, let $H(n,k,t)$ be the number of binary trees $T \in T(n)$ having exactly k nodes in the stack after t units of time. An inspection of Figure 1 shows that this number is the product of

(a) the number of paths from $(1,1)$ to (t,k), which is

$$\frac{k}{t}\binom{t}{(t+k)/2}, \quad \text{and}$$

(b) the number of paths from (t,k) to $(2n-1,1)$, which is

$$\frac{k}{2n-t}\binom{2n-t}{n-(t+k)/2}.$$

These enumeration results of the number of paths are well-known ([3]). Hence

$$H(n,k,t) = \frac{k^2}{t(2n-t)}\binom{t}{(t+k)/2}\binom{2n-t}{n-(t+k)/2}. \tag{1}$$

Obviously, we have the conditions $k \leqq t \leqq 2n-1$ and $(k+t) \equiv 0 \bmod 2$. Now, let $|T(n)| = t(n)$. It is well-known ([6]) that

$$t(n) = \frac{1}{n}\binom{2n-2}{n-1}. \tag{2}$$

Considering all binary trees $T \in T(n)$ equally likely the quotient $p(n,k)$

$=H(n,k,t)/t(n)$ is the probability that there are k nodes in the stack after t units of time during postorder-traversing of a binary tree $T \in T(n)$. Therefore, the sth moment is given by

$$R_s(n,t) = t(n)^{-1} \sum_{k=1}^{t} k^s H(n,k,t) . \tag{3}$$

Supposing $b \in \mathbf{Z}$ and using the explicit expression for $H(n,k,t)$ given in (1) we obtain by a simple calculation

$$\begin{aligned} &R_s(n,2t+b) \\ &= 2\frac{2n-1}{(2t+b)(2n-2t-b)} \sum_{k\geqq 0} (2k+b)^{s+2} \binom{2t+b}{t-k}\binom{2n-2t-b}{n-t+k}\Big/\binom{2n}{n} \end{aligned} \tag{4}$$

where $b \in \{0,1\}$. In order to compute $R_s(n,t)$ as a function of the proportion of the units of time t to the whole number of units of time $2n$ we have to compute an asymptotic equivalent for the sum

$$h_{b,s}(n,t) = \sum_{k\geqq 0} k^s \binom{2t+b}{t-k}\binom{2n-2t-b}{n-t+k}\Big/\binom{2n}{n}, \quad s\geqq 0 \tag{5}$$

because

$$R_s(n,2t+b) = 2\frac{2n-1}{(2t+b)(2n-2t-b)} \sum_{\mu=0}^{s+2} \binom{s+2}{\mu} 2^\mu b^{s+2-\mu} h_{b,\mu}(n,t) . \tag{6}$$

In [5] a method is given for the computation of a closed expression for $h_{0,3}(n,t)$; this method works also for $h_{0,s}(n,t)$ where s is odd, but not for even s. In this paper we shall give another procedure for the computation of an asymptotic equivalent to $h_{b,s}(n,t)$ effective for all $s \geqq 0$.

Let $\varrho_b = (t+b/2)/n$, $x = (k+b/2)/(\varrho_b n)$ and for $i \geqq 1$

$$a_i = \frac{\varrho_b}{i(2i-1)}\left[1+\left(\frac{\varrho_b}{1-\varrho_b}\right)^{2i-1}\right], \quad b_i = \frac{1}{i}\left[1+\left(\frac{\varrho_b}{1-\varrho_b}\right)^{2i}\right],$$

$$c_0 = \frac{1}{8}\left[1-\frac{1}{\varrho_b(1-\varrho_b)}\right], \quad c_i = \frac{1}{6\varrho_b}\left[1+\left(\frac{\varrho_b}{1-\varrho_b}\right)^{2i+1}\right].$$

For $\varrho_b = \text{const.}$ we have by Stirling's approximation

$$\begin{aligned} \binom{2t+b}{t-k}\binom{2n-2t-b}{n-t+k}\Big/\binom{2n}{n} &= (\pi n \varrho_b(1-\varrho_b))^{-\frac{1}{2}} \exp(-n(a_1x^2+a_2x^4+\ldots)+ \\ &+\tfrac{1}{2}(b_1x^2+b_2x^4+\ldots)+\frac{1}{n}(c_0+c_1x^2+c_2x^4+\ldots)+O(x^2n^{-3})) \end{aligned}$$

when $-1<x<+1$, and $\binom{2t+b}{t-k}\binom{2n-2t-b}{n-t+k}\Big/\binom{2n}{n} = O(n^{-\frac{1}{2}}\exp(-n^{2\varepsilon}))$

when $k \geqq \sqrt{\{\varrho_b(1-\varrho_b)\}}\, n^{\varepsilon+\frac{1}{2}} - b/2$, for all fixed $\varepsilon > 0$. Therefore the sum of all terms for $k \geqq \sqrt{\{\varrho_b(1-\varrho_b)\}}\, n^{\varepsilon+\frac{1}{2}} - b/2$ in (5) is negligible, being $O(n^{-m})$ for all $m > 0$. Hence we may take $x = O(n^{-\frac{1}{2}+\varepsilon})$ in (5). Therefore

$$\binom{2t+b}{t-k}\binom{2n-2t-b}{n-t+k}\Big/\binom{2n}{n} = (\pi n \varrho_b(1-\varrho_b))^{-\frac{1}{2}} \exp\left(-n(a_1x^2+a_2x^4)+\tfrac{1}{2}b_1x^2+c_0/n+O(n^{-2+\varepsilon})\right).$$

Using the definition of x, ϱ_b and the above explicit expressions for a_1, a_2, b_1, c_0 we obtain by an elementary computation

$$(7)\quad \binom{2n+b}{t-k}\binom{2n-2t-b}{n-t+k}\Big/\binom{2n}{n} = (\pi n \varrho_b(1-\varrho_b))^{-\frac{1}{2}} \exp\left(-\frac{(k+b/2)^2}{n\varrho_b(1-\varrho_b)}+A_{n,k,b}\right)$$

where

$$A_{n,k,b} = \frac{1}{8n} - \frac{1}{8n\varrho_b(1-\varrho_b)} + \frac{(1-\varrho_b)^2+\varrho_b^2}{2n^2\varrho_b^2(1-\varrho_b)^2}(k+\tfrac{1}{2}b)^2 - \\ - \frac{(1-\varrho_b)^3+\varrho_b^3}{6n^3\varrho_b^3(1-\varrho_b)^3}(k+\tfrac{1}{2}b)^4 + O(n^{-1.5+\varepsilon}).$$

We now consider the function

$$g_{l,b}(n,t) = \sum_{k \geqq 1} k^l \exp\left(-k^2/(n\varrho_b(1-\varrho_b))\right), \qquad l \text{ fixed}.$$

Again the terms for $k \geqq \sqrt{\{\varrho_b(1-\varrho_b)\}}\, n^{\varepsilon+\frac{1}{2}} - b/2$ are negligible. Hence, we can use (7) to express $h_{b,s}(n,t)$ in terms of $g_{l,b}(n,t)$:

$$(8)\quad h_{b,s}(n,t) = (\pi n \varrho_b(1-\varrho_b))^{-\frac{1}{2}} \left[\left(1 - \frac{\varrho_b^2-\varrho_b+2b^2+1}{8n\varrho_b(1-\varrho_b)}\right) g_{s,b}(n,t) - \right. \\ - \frac{b}{n\varrho_b(1-\varrho_b)} g_{s+1,b}(n,t) + \frac{2\varrho_b^2-2\varrho_b+b^2+1}{2n^2\varrho_b^2(1-\varrho_b)^2} g_{s+2,b}(n,t) - \\ \left. - \frac{3\varrho_b^2-3\varrho_b+1}{6n^3\varrho_b^3(1-\varrho_b)^3} g_{s+4,b}(n,t) + O(g_{s,b}(n,t)n^{-1.5+\varepsilon})\right].$$

Now we compute the asymptotic behavior of $g_{s,b}(n,t)$. With the wellknown formula

$$\exp(-x) = (2\pi i)^{-1} \int_{c-i\infty}^{c+i\infty} \Gamma(z) x^{-z}\, dz, \qquad x > 0, \qquad c > 0,$$

where $\Gamma(z)$ is the complete gamma function, we find

$$g_{l,b}(n,t) = (2\pi i)^{-1} \int_{c-i\infty}^{c+i\infty} \Gamma(z) n^z \varrho_b^z (1-\varrho_b)^z \zeta(2z-l)\, dz$$

with $c > (l+1)/2$. Here, $\zeta(z)$ is the Riemann zeta function. It can be shown by a well-known method that we can shift the integration line to the left as far as we please if we only take the residues into account. There are simple poles at $z = (l+1)/2$ and possibly at $z = -k$, $k \in \mathbb{N}_0$. A computation of these residues leads to

$$g_{l,b}(n,t) = \tfrac{1}{2}\Gamma(\tfrac{1}{2}(l+1))[n\varrho_b(1-\varrho_b)]^{(l+1)/2} + \zeta(-l) - \frac{\zeta(-2-l)}{n\varrho_b(1-\varrho_b)} + O(n^{-2}) .$$

Therefore we get from (8) by an elementary calculation for all $\varepsilon > 0$:

$$\begin{aligned} h_{b,s}(n,t) = {} & \tfrac{1}{2}\pi^{-\frac{1}{2}}[p_1 y^{s/2} + p_2 y^{(s-1)/2} + p_3 y^{(s-2)/2} + p_4 y^{-\frac{1}{2}} + \\ & + O(n^{(s-3)/2+\varepsilon})] \end{aligned}$$

where $y = n\varrho_b(1-\varrho_b)$ and:

$$\begin{aligned} p_1 &= \Gamma(\tfrac{1}{2}(s+1)), \\ p_2 &= -b\Gamma(\tfrac{1}{2}(s+2)), \\ p_3 &= -\tfrac{1}{8}\Gamma(\tfrac{1}{2}(s+1))[\varrho_b^2 - \varrho_b + 2b^2 + 1] + \tfrac{1}{2}\Gamma(\tfrac{1}{2}(s+3))[2\varrho_b^2 - 2\varrho_b + b^2 + 1] - \\ & \quad -\tfrac{1}{6}\Gamma(\tfrac{1}{2}(s+5))[3\varrho_b^2 - 3\varrho_b + 1], \\ p_4 &= 2\zeta(-s) . \end{aligned}$$

Returning to (6) we obtain with this approximation the following

LEMMA. *Let $b \in \{0,1\}$ and $\varrho_b = (t + \frac{1}{2}b)/n$. The* sth *moment is given for all $\varepsilon > 0$ by*

$$(a) \quad R_1(n, 2t+b) = \sqrt{(n/\pi)}\left[4\sqrt{\{\varrho_b(1-\varrho_b)\}} - \frac{5\varrho_b^2 - 5\varrho_b + 1}{2n\sqrt{\{\varrho_b(1-\varrho_b)\}}} + O(n^{-3/2+\varepsilon})\right]$$

$$(b) \quad R_2(n, 2t+b) = n\left[6\varrho_b(1-\varrho_b) - \frac{1}{n}(9\varrho_b^2 - 9\varrho_b + 2) + O(n^{-3/2+\varepsilon})\right]$$

$$\begin{aligned} (c) \quad R_s(n, 2t+b) = {} & \pi^{-\frac{1}{2}} 2^{s+1}\Gamma(\tfrac{1}{2}(s+3))\,[n\varrho_b(1-\varrho_b)]^{s/2} + \\ & + O(n^{(s-1)/2}) \quad \textit{for } s \geqq 3 . \end{aligned}$$

Since $R_1(n,t)$ is the average number of nodes in the stack after t units of time during postorder-traversing of a tree $T \in T(n)$ and the variance $\sigma^2(n)$ is given by $R_2(n,t) - R_1^2(n,t)$ we get with the preceding Lemma the following

THEOREM. *Assuming that all binary trees with n leaves are equally likely the average number of nodes $R_1(n,t)$ stored in the stack after t units of time during postorder-traversing of $T \in T(n)$ is given for all $\varepsilon > 0$ by*

$$R_1(n,t) = 4\sqrt{\{n\varrho(1-\varrho)/\pi\}} - \frac{5\varrho^2 - 5\varrho + 1}{2\sqrt{\{\pi n\varrho(1-\varrho)\}}} + O(n^{-1+\varepsilon})$$

where $\varrho = t/2n = const.$ *For all* $\varepsilon > 0$ *the variance is*

$$\sigma^2(n,t) = (6-16/\pi)\varrho(1-\varrho)n + (9-20/\pi)\varrho(1-\varrho) - (2-4/\pi) + O(n^{-\frac{1}{2}+\varepsilon}).$$

Figure 2 shows the graph of $R_1(n,t)$ and $\sigma(n,t)$ as functions of the proportion of the units of time t to the whole number of units of time $2n$ needed to traverse a tree $T \in T(n)$.

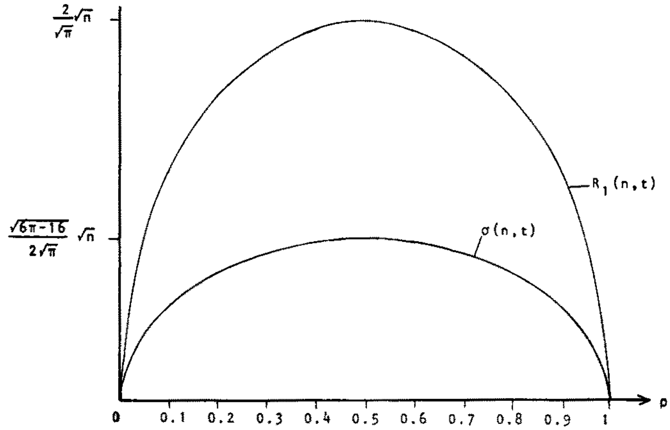


Figure 2. The average stack size and the standard deviation as functions of $\varrho = t/2n$.

REFERENCES

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, INC, New York, 1970.
2. T. M. Apostol, *Introduction to Analytic Number Theory*, Springer Verlag, New York, 1976.
3. L. Carlitz, D. P. Roselle and R. A. Scoville, *Some remarks on ballot-type sequences of positive integers*, J. Comb. Theory, Ser. A 11 (1971), 258–271.
4. N. G. deBruijn, D. E. Knuth and S. O. Rice, *The Average Height of Planted Plane Trees*, R. C. Read (Ed.), Graph Theory and Computing, New York, London, Ac. Press (1972), 15–22.
5. R. Kemp, *On The Average Stack Size of Regularly Distributed Binary Trees*, H. A. Maurer (Ed.), Lect. Notes in Comp. Sci. 71 (1979), 340–355.
6. D. E. Knuth, *The Art of Computer Programming*, Vol. 1, 2nd ed., Addison–Wesley, Reading, 1973.
7. G. Kreweras, *Sur les éventails de segments*, Cahiers du B.U.R.O. 15, Paris, (1970), 1–41.
8. J. Riordan, *Combinatorial Identities*, Wiley, New York, 1968.

UNIVERSITÄT DES SAARLANDES
FACHBEREICH 10
D-6600 SAARBRÜCKEN
WEST-DEUTSCHLAND