The Art of Computer Programming

Explicit Solutions

Keaun Moughari

November 1, 2022

§ 1.2.4

42. re: Lower-bound of internal path length of extended binary trees (p. 400)

(a) Prove that $\sum_{k=1}^{n} a_k = na_n - \sum_{k=1}^{n-1} k(a_{k+1} - a_k), \quad \text{if } n > 0.$

Solution:

$$\begin{aligned}
\sum_{k=1}^{n} a_k &= na_n - \sum_{k=1}^{n-1} k(a_{k+1} - a_k) \\
&= na_n + \sum_{k=1}^{n-1} ka_k - \sum_{k=1}^{n-1} ka_{k+1} \\
&= na_n + (a_1 + 2a_2 + \cdots + (n-1)a_{n-1}) \\
&\quad - (a_2 + 2a_3 + \cdots + (n-2)a_{n-1} + (n-1)a_n) \\
&= na_n - (n-1)a_n + (a_1 + 2a_2 + \cdots + (n-1)a_{n-1}) \\
&\quad - (a_2 + 2a_3 + \cdots + (n-2)a_{n-1}) \\
&= a_n + \sum_{k=1}^{n-1} ka_k - \sum_{k=1}^{n-2} ka_{k+1} \\
&= a_n + \sum_{k=0}^{n-2} (k+1)a_{k+1} - \sum_{k=1}^{n-2} ka_{k+1} \\
&= a_n + a_1 + \sum_{k=1}^{n-2} (k+1)a_{k+1} - \sum_{k=1}^{n-2} ka_{k+1} \\
&= a_n + a_1 + \sum_{k=1}^{n-2} (k+1)a_{k+1} - ka_{k+1} \\
&= a_n + a_1 + \sum_{k=1}^{n-2} a_{k+1} \\
&= a_n + a_1 + \sum_{k=2}^{n-1} a_k \\
&= \sum_{k=1}^{n} a_k
\end{aligned}$$

(b) The preceding formula is useful for evaluating certain sums involving the floor function. Prove that, if b is an integer ≥ 2,

$$\sum_{k=1}^{n} \lfloor \log_b k \rfloor = (n+1)\lfloor \log_b n \rfloor - (b^{\lfloor \log_b n \rfloor + 1} - b)/(b-1).$$

Aside:

(1) $f = \lfloor \log_b n \rfloor; b \geq 2$

$$\begin{aligned}
S_{n-1} &= \sum_{k=1}^{f-1} b^k \\
&= b - b^{\lfloor \log_b n \rfloor} + \sum_{k=1}^{f-1} b^{k+1} \\
&= b - b^{\lfloor \log_b n \rfloor} + b \sum_{k=1}^{f-1} b^k \\
&= b - b^{\lfloor \log_b n \rfloor} + bS_{n-1} \\
(1-b)S_{n-1} &= b - b^{\lfloor \log_b n \rfloor} \\
\implies S_{n-1} &= \frac{b^{\lfloor \log_b n \rfloor} - b}{b-1}
\end{aligned}$$

(2) $$\sum_{1}^{f-1} = \sum_{j}^{f-1} + \sum_{1}^{j-1}$$

$$\implies \sum_{j}^{f-1} = \sum_{1}^{f-1} - \sum_{1}^{j-1}$$

(3) $f = \lfloor \log_b n \rfloor; b \geq 2$

$$\begin{aligned}
S_n &= \sum_{k=1}^{n-1} k \; [k+1 \text{ is a power of } b] \\
&= \sum_{k=1}^{f} (b^k - 1) \\
&= \sum_{k=1}^{f} b^k - \sum_{k=1}^{f} 1 \\
&= \sum_{k=1}^{f} b^k - f \\
&= \left[fb^f - \sum_{k=1}^{f-1} k(b^{k+1} - b^k) \right] - f \qquad (a) \\
&= f(b^f - 1) - \sum_{k=1}^{f-1} k(b^{k+1} - b^k) \\
&= f(b^f - 1) - (b-1) \sum_{k=1}^{f-1} kb^k \\
&= f(b^f - 1) - (b-1) \sum_{j=1}^{f-1} \sum_{k=j}^{f-1} b^k
\end{aligned}$$

$$
\begin{aligned}
&= f(b^f - 1) - (b-1)\sum_{j=1}^{f-1}\left(\sum_{k=1}^{f-1} b^k - \sum_{k=1}^{j-1} b^k\right) && \textit{(2)}\\
&= f(b^f - 1) - (b-1)\sum_{j=1}^{f-1}\left(\frac{b^f - b}{b-1} - \frac{b^j - b}{b-1}\right)\\
&= f(b^f - 1) - (b-1)\sum_{j=1}^{f-1}\left(\frac{b^f}{b-1} - \frac{b^j}{b-1}\right)\\
&= f(b^f - 1) - (b-1)\left[\frac{1}{b-1}\left(\sum_{j=1}^{f-1} b^f - \sum_{j=1}^{f-1} b^j\right)\right]\\
&= f(b^f - 1) - \left(\sum_{j=1}^{f-1} b^f - \sum_{j=1}^{f-1} b^j\right)\\
&= f(b^f - 1) - \left(b^f(f-1) - \sum_{j=1}^{f-1} b^j\right)\\
&= f(b^f - 1) - \left(b^f(f-1) - \left(\frac{b^f - b}{b-1}\right)\right)\\
&= f(b^f - 1) - b^f(f-1) + \left(\frac{b^f - b}{b-1}\right)\\
&= fb^f - f - fb^f + b^f + \left(\frac{b^f - b}{b-1}\right)\\
&= b^f - f + \left(\frac{b^f - b}{b-1}\right)\\
(b-1)S_n &= (b-1)\left(b^f - f\right) + \left(b^f - b\right)\\
&= b^{f+1} - fb + f - b^f + b^f - b\\
&= b^{f+1} - b - fb + f\\
&= b^{f+1} - b - f(b-1)\\
S_n &= \left(\frac{b^{f+1} - b}{b-1}\right) - f\\
\implies S_n &= \frac{b^{\lfloor \log_b n \rfloor + 1} - b}{b-1} - \lfloor \log_b n \rfloor
\end{aligned}
$$

e.o.a.

Solution:

$$
\begin{aligned}
\sum_{k=1}^{n} \lfloor \log_b k \rfloor &= n\lfloor \log_b n \rfloor - S_n\\
&= n\lfloor \log_b n \rfloor - \left[\frac{b^{\lfloor \log_b n \rfloor + 1} - b}{b-1} - \lfloor \log_b n \rfloor\right] && (3)\\
&= n\lfloor \log_b n \rfloor - \frac{b^{\lfloor \log_b n \rfloor + 1} + b}{b-1} + \lfloor \log_b n \rfloor
\end{aligned}
$$

$$= (n+1)\lfloor \log_b n \rfloor - \frac{b^{\lfloor \log_b n \rfloor + 1} + b}{b-1}$$

§ 2.3.4.5

5. re: Generating function of BT with n nodes and internal path length p (p. 404)

Let

$$B(w,z) = \sum_{n,p \geq 0} b_{np} w^p z^n,$$

where b_{np} is the number of binary trees with n nodes and internal path length p.

(a) Find a functional relation that characterizes $B(w,z)$, generalizing $zB(z)^2 = B(z) - 1$.

(b) Use the result of (a) to determine the average internal path length of a binary tree with n nodes, assuming that each of the $\frac{1}{n+1}\binom{2n}{n}$ trees is equally probable.

(c) Find the asymptotic value of this quantity.

§ 6.2.1

25. re: Relation between internal and external nodes in extended BT (p. 425)

Suppose that a binary tree has a_k internal nodes and b_k external nodes on level k, for $k = 0, 1, \ldots$ (The root is at level zero.) Thus in Fig. 8 we have $(a_0, a_1, \ldots, a_5) = (1, 2, 4, 4, 1, 0)$ and $(b_0, b_1, \ldots, b_5) = (0, 0, 0, 4, 7, 2)$.

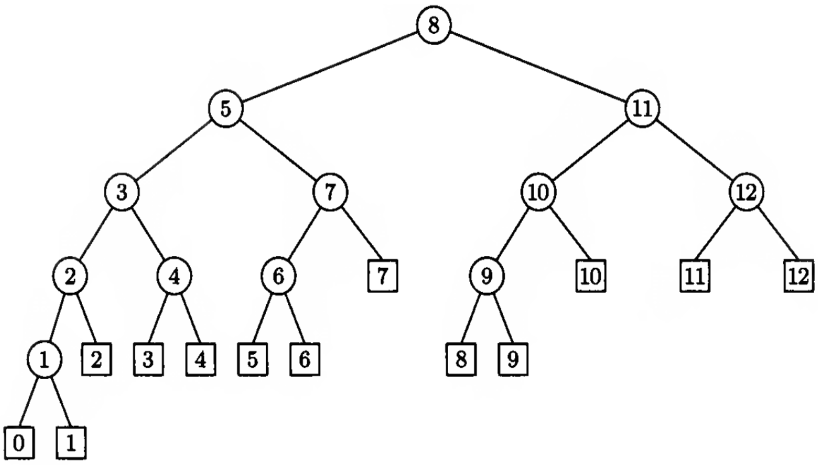

Fig. 8. The Fibonacci tree of order 6.

(a) Show that a simple algebraic relationship holds between the generating functions $A(z) = \sum_k a_k z^k$ and $B(z) = \sum_k b_k z^k$.

Solution:

The number of internal nodes at each level k in a binary tree is twice the number of internal nodes at the previous level minus the number of external nodes at level k (for $k > 0$, as level 0 has exactly 1 internal node). With that in mind, the generating function, $A(z)$, can be modified to represent the total possible number of internal nodes at each level. Then, to calculate the number of internal nodes, the number of external nodes can be subtracted from this modified representation.

$$\begin{aligned}
A(z) + B(z) &= (a_0 + b_0) + (a_1 + b_1)z + \ldots + (a_n + b_n)z^n \\
2A(z) &= (a_0 + a_0) + (a_1 + a_1)z + \ldots + (a_n + a_n)z^n \\
2zA(z) &= 0 + 2a_0 z + 2a_1 z^2 + \ldots + 2a_{n-1}z^n \\
1 + 2zA(z) &= 1 + 2a_0 z + 2a_1 z^2 + \ldots + 2a_{n-1}z^n \\
&= A(z) + B(z)
\end{aligned}$$

$$\begin{aligned}
a_0 + b_0 &= 1 & \implies \quad a_0 &= 1 - b_0 = 1 \\
a_1 + b_1 &= 2a_0 & \implies \quad a_1 &= 2a_0 - b_1 = 2 \\
a_2 + b_2 &= 2a_1 & \implies \quad a_2 &= 2a_1 - b_2 = 4 \\
a_3 + b_3 &= 2a_2 & \implies \quad a_3 &= 2a_2 - b_3 = 4 \\
a_4 + b_4 &= 2a_3 & \implies \quad a_4 &= 2a_3 - b_4 = 1 \\
a_5 + b_5 &= 2a_4 & \implies \quad a_5 &= 2a_4 - b_5 = 0
\end{aligned}$$

As seen above, the total number of nodes at each level can be expressed exclusively in terms of $A(z)$ and an algebraic relationship between both representations exists; namely, $A(z) = 1 + 2zA(z) - B(z)$.

(b) The probability distribution for a successful search in a binary tree has the generating function $g(z) = zA(z)/N$, and for an unsuccessful search the generating function is $h(z) = B(z)/(N + 1)$. (Thus in the text's notation we have $C_N = \text{mean}(g), C_N' = \text{mean}(h)$, and Eq. 2 ($C_N = (1+\frac{1}{N})C_N' - 1$) gives a relation between these quantities.) Find a relation between $\text{var}(g)$ and $\text{var}(h)$.