Letter to the Editor

A Note on Binomial Recurrences Arising in the Analysis of Algorithms

Helmut Prodinger*
Department of Algebra&Discrete Mathematics
Technical University of Vienna
Vienna A-1040
Austria

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Chang-Biau Yang in "Reducing Conflict Resolution Time for Solving Graph Problems in Broadcast Communications" *Information Processing Letters*, **40**, 295–302, 1992 analyzed the following recurrence (we replace the authors notation C_k^n for a binomial coefficient by the more common $\binom{n}{k}$)

$$g_n = 1 + \sum_{k=0}^{n} 2^{1-n} \binom{n}{k} g_k \tag{1}$$

for $n \geq 2$ and $g_0 = g_1 = 1$.

The author presents a lengthy derivation for the following bounds on g_n:

$$\frac{8}{3} \leq \frac{g_n}{n} \leq 3$$

However, such recursions were intensively studied since the appearance of Knuth's books [13] in the context of *tries*, and have gained more attention in the past few years in the context of *conflict resolution protocols* [14], [17], analysis of extendible hashing [2], [19], and so forth.

The exact answer to (1) is (cf. [13], [17])

$$g_n = 1 + \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{2k-2}{1-2^{1-k}} \ . \tag{2}$$

Much more can be said. The asymptotic expansion of (2) leads to (cf. [14], [13], [17], [19])

$$g_n = \frac{2}{\log 2} n + n\delta(\log_2 n) + \textit{smaller order terms} \ , \tag{3}$$

*This work is supported by Fonds zur Förderung der Wissenschaftlichen Forschung Project P7497-TEC

†This research was supported by the NSF Grants INT-8912631, CCR-8900305 (its extension CCR-9201078) and NCR-9206315, in part by AFOSR grant 90-0107, and in part by grant R01 LM05118 from the National Library of Medicine.

where $\delta(x)$ is a periodic function of period 1, small amplitude, mean 0 and known Fourier coefficients. Note that $2/\log 2 = 2.8853\ldots$.

Computing the variance of the conflict resolution session (the quantity analyzed by C. Yang) is much harder, but it can be done using a general methodology explained briefly below. The difficulty arises from a new sort of recurrence that must be solved. This recurrence (for the second factorial moment of g_n) is of the following type

$$h_n = 2ng_n + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}g_k g_{n-k} - n(n+1) + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}h_k \ . \tag{4}$$

The above recurrence is difficult to solve mainly due to the appearance of g_n (cf. (2)) in the second term above. But, even the first term is not quite trivial to handle. In Kirschenhofer, Prodinger and Szpankowski [11, 12] asymptotic expansions of h_n and the variance (i.e., $h_n + g_n - g_n^2$) are presented. Such an analysis involves non trivial identities from the theory of modular functions and Ramanujan's identities (cf. [10]). The interested reader is referred to the new book of Hosam Mahmoud [15] which treats this subject in a more detailed way.

In this short note we only indicate how (2) and (3) can be obtained. Consider a general recurrence equation of the following type

$$x_n = a_n + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}x_k \qquad n \geq 2 \ , \tag{5}$$

given x_0 and x_1, where a_n is any sequence. (For (1) we need to assume that $a_n = 1$ and $x_0 = x_1 = 1$.) In [13, 17, 18] it is shown that such a recurrence possesses the following solution

$$x_n = x_0 + \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{k\bar{x}_1 + \hat{a}_k - \bar{x}_0}{1 - 2^{1-k}} \tag{6}$$

where $\bar{x}_0 = a_0 + x_0$ and $\bar{x}_1 = a_1 + x_0$. The inverse relationship $\hat{a}_n$ is defined as follows (cf. Knuth [13])

$$\hat{a}_n = \sum_{k=0}^{n}\binom{n}{k}(-1)^k a_k \ . \tag{7}$$

For example, if $a_n = 1$, then $\hat{a}_n = \delta_{n,0}$. Using this, we can easily obtain solution (2) of the recurrence (1).

The asymptotic expansion of the alternating sum (6) requires some tools from complex analysis [6]. Consider a general alternating sum of the following type

$$x_n = \sum_{k=2}^{n}(-1)^k\binom{n}{k}f_k \ , \tag{8}$$

An asymptotic expansion of the above sum can be either obtained by Rice's method as popularized by Knuth [13] and Flajolet and Sedgewick [4] (see also [9]) or by the Mellin-like transform as shown by Knuth [13] and generalized by Szpankowski [18]. We discuss below only the Mellin-like approach. In Szpankowski [18] it is proved that the alternating sum (8) can be evaluated by the following complex integral

$$x_n = \frac{1}{2\pi i}\int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z) n^{-z} f(-z) dz + \textit{smaller order terms} \;, \tag{9}$$

where $\Gamma(z)$ is the gamma function [6], $f(z)$ is an analytical continuation of f_k, that is, $f(k) = f_k$. The above formula holds under some mild conditions on the behaviour of $f(z)$ at infinity. In passing, we note that the integral (9) is usually *very easy* to evaluate asymptotically by Cauchy's residue theorem [6].

In the case of the recurrence (1) with the explicit solution (2), we apply (9) to obtain

$$g_n \sim -\frac{1}{2\pi i}\int_{-3/2-i\infty}^{-3/2+i\infty} \frac{(2z+2)\Gamma(z)n^{-z}}{1-2^{1+z}}\, dz \;. \tag{10}$$

Noting that $z_0 = -1$ is a double pole of the function under the integral and computing the residue at z_0, we obtain the first term in (3). The second term (oscillating function $\delta(\log_2 n)$) is derived from the infinite number of poles of the form $z_k = -1 + 2\pi k i/\log 2$ of the denominator. Details can be found in the references already mentioned above.

In passing, we note that in fact much more was done in this area. We refer to seminal works Knuth and Flajolet, and some others, like (in alphabetical order) Jacquet, Kirschenhofer, Prodinger, Régnier, Schmid, Sedgewick and Szpankowski.

Added in the Proof. Yet another binomial recurrence was consider by Ressler in his recent paper "Random List Permutations in Place", *Information Processing Letters*, **43**, 271-275, 1992. Namely, for $n \geq 0$

$$T_{n+1} = n + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}T_k \tag{11}$$

with $T_0 = 0$, and Ressler proved that $T_n \leq n\log_2 n$. This recurrence does not fall exactly under our general scheme (5). In fact, it is harder to solve, and it appears in the analysis of another digital tree, namely, *digital search tree* (cf. [3], [4], [5], [13], [21]).

More generally, as in [21] consider the following recurrence for $n \geq 1$

$$x_{n+1} = a_{n+1} + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}x_k \tag{12}$$

with $x_0 = 0$. It is proved in [21] that the above recurrence has the following solution

$$x_n = -\sum_{k=2}^{n} (-1)^k \binom{n}{k} \hat{x}_{k-2} \tag{14}$$

where

$$\hat{x}_n = Q_n \sum_{i=1}^{n+1} (\hat{a}_i - \hat{a}_{i+1} - a_1)/Q_{i-1} \tag{15}$$

with $Q_n = \prod_{k=1}^{n}(1 - 2^{-k})$. Note that the solution (14) has the form of the alternating sum as in (8), hence the methodology discussed above can be directly applied to obtain an asymptotic expansion of (14).

In particular, for the recurrence (11) studied by Ressler we can easily prove that $T_n = \sum_{k=2}^{n}(-1)^k \binom{n}{k} Q_{k-2}$, and

$$ET_n = n\log_2 n + n\left(\frac{\gamma - 1}{\log 2} + \frac{1}{2} - \alpha + \delta_1(\log_2 n)\right) + O(\log n) \ , \tag{16}$$

where $\gamma = 0.57721\ldots$ is Euler's constant and $\alpha = \sum_{n\geq 1} 1/(2^n - 1) = 1.60669\ldots$, and $\delta_1(x)$ is a continuous periodic function of period 1, mean 0 and very small amplitude ($< 10^{-6}$). For more details, the reader is referred to the literature cited above.

References

[1] J. Capetanakis, Tree Algorithms for Packet Broadcast Channels, *IEEE Trans. on Inf. Theory*, 25, 505-515 (1979).

[2] P. Flajolet, On the Performance Evaluation of Extendible Hashing and Trie Searching, *Acta Informatica* 20, 345-369 (1983).

[3] P. Flajolet, Analytic Analysis of Algorithms, *Proc. ICALP 92*, Vienna, Lecture Notes in Computer Science, vol. 623, Springer Verlag, 186-210 (1992).

[4] P. Flajolet and R. Sedgewick, Digital Search Trees Revisited, *SIAM J. Computing*, 15, 748-767 (1986).

[5] P. Flajolet and B. Richmond, Generalized Digital Trees and Their Difference-Differential Equations, *Random Structures & Algorithms*, 3, 305-320 (1992).

[6] P. Henrici, *Applied and Computational Complex Analysis*, Vol 2., John Wiley&Sons, New York 1977.

[7] P. Jacquet and M. Régnier, Limiting Distributions for Trie Parameters, *Lecture Notes in Computer Science*, 214, 196-210 (1986).

[8] P. Jacquet and W. Szpankowski, Analysis of Digital Tries with Markovian Dependency, *IEEE Transactions on Information Theory*, 37, 1470-1475 (1991).

[9] P. Kirschenhofer and H. Prodinger, Further Results on Digital Search Trees, *Theoretical Computer Science*, 58, 143-154 (1988).

[10] P. Kirschenhofer and H. Prodinger, On Some Applications of Formulae of Ramanujan in the Analysis of Algorithms, Mathematika, 38, 14-33 (1991).

[11] P. Kirschenhofer, H. Prodinger and W. Szpankowski, On the Variance of the External Path length in a Symmetric Digital Tries, *Discrete Applied Mathematics*, 25, 129-143 (1989).

[12] P. Kirschenhofer, H. Prodinger and W. Szpankowski, On the balance Property of Patricia Tries: External Path Length Viewpoint, *Theoretical Computer Science*, 68, 1-17 (1989).

[13] D.E. Knuth, *The Art of Computer Programming. Vol. 3*, Addison-Wesley, Reading MA 1973.

[14] P. Mathys and P. Flajolet, Q-ary Collision Resolution Algorithms in Random-Access Systems with Free or Blocked Channels Access, *IEEE Trans. Information Theory*, 31, 217-243 (1985).

[15] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York 1992.

[16] U. Schmid, On a Tree Collision Resolution Algorithm in Presence of Capture, *RAIRO, Theoretical Informatics*, 26, 163-197 (1992).

[17] W. Szpankowski, On a Recurrence Equation Arising in the Analysis of Conflict Resolution Algorithms, *Stochastic Models*, 3, 89-114 (1987).

[18] W. Szpankowski, The Evaluation of an Alternating Sum with Applications to the Analysis of Some Data Structures, *Information Processing Letters*, 28, 13-19 (1988).

[19] W. Szpankowski, Some Results on V-ary Asymmetric Tries, *Journal of Algorithms*, 9, 224-244 (1988).

[20] W. Szpankowski, Patricia Tries Again Revisited, *Journal of the ACM,* 37, 691-711 (1991).

[21] W. Szpankowski, A Characterization of Digital Search Trees from the Successful Search Viewpoint, *Theoretical Computer Science*, 85, 117-134 (1991).