JOURNAL OF COMBINATORICS,
INFORMATION & SYSTEM SCIENCES

Vol. 5, No. 4, 287-304 (1980)

The Average Height of a Stack where Three Operations are Allowed and Some Related Problems

HELMUT PRODINGER

Institut für Mathematische Logik und Formale Sprachen,
Technische Universität Wien, Gußhausstraße 27-29, 1040 Wien, Austria

The average stack size after n units of time of a stack where the three operations insertion, deletion, interrogation are allowed, is

$$\sqrt{\frac{\pi n}{3}}-\frac{3}{2}+0\left(\frac{\log n}{n^{1/2-\epsilon}}\right),\ n\to\infty,\ \text{for all } \epsilon>0.$$

Some related results are given.

1. INTRODUCTION AND PRELIMINARIES

De Bruijn, Knuth, Rice [2] have determined the asymptotic behaviour of the average height of planted plane trees; this result is easily reformulated in terms of, say, random walks or paths or Dyck-words or ballot sequences or binary trees.

If a path (or closed path) of length n means a sequence $(0, s_0)$, $(1, s_1), \ldots, (n, s_n)$ with $s_0 = s_n = 0$, $s_i \geqslant 0$, $|s_{i+1} - s_i| = 1$ and the height of this path is $\max\limits_{0\leqslant i\leqslant n} s_i$, then the average height of a path of length $2n$ is

$$\sqrt{\pi n}-\frac{3}{2}+0\left(\frac{\log n}{n^{1/2-\epsilon}}\right),\ n\to\infty,\ \text{for all } \epsilon>0.$$

Kemp [8] has computed further terms of this asymptotic series and also higher moments.

In sections 2 to 4 an analogous problem is considered, namely the following: The condition $|s_{i+1} - s_i| = 1$ is replaced by $s_{i+1} - s_i \in \{-1, 0, 1\}$. This situation can be reformulated in terms of the history of a stack (see Flajolet [5]), where the 3 operations insertion (+1), deletion (−1) and interrogation (0) are allowed. The result is: The average height of such a path ("of 3 symbols") of length n is

$$\sqrt{\frac{\pi n}{3}}-\frac{3}{2}+0\left(\frac{\log n}{n^{1/2-\epsilon}}\right),\ n\to\infty,\quad \text{for all } \epsilon>0.$$

The result of [2] is obtained by approximating the binomial coefficients by means of the Stirling's formula by the normal distribution.

Here the situation is more complicated: the so-called trinomial coefficients appear (i.e. the coefficients of $(1 + x + x^2)^n$), and these must be approximated by the normal distribution (Section 3).

In Section 4 the restriction $s_i \geqslant 0$ is dropped (the classical case $|s_{i+1} - s_i| = 1$ is again considered). The average height $(\max_{0 \leqslant i \leqslant n} |s_i|)$ of such a path of length $2n$ is

$$(\log 2) \cdot \sqrt{\pi n} - \frac{1}{2} + 0\left(\frac{1}{\sqrt{n}}\right), \quad n \to \infty.$$

R. Kemp [8] has computed the average stack size after t units of time during postorder-traversing of a binary tree with n leaves. His result is

$$\frac{2}{\sqrt{\pi n}} \sqrt{t(2n - t)} \cdot \left\{1 + 0\left(\frac{1}{t}\right) + 0\left(\frac{1}{n - t}\right)\right\}, \quad t \to \infty, \ n - t \to \infty.$$

A reformulation of this result in terms of paths gives: The average value s_t of a path ($s_0 = s_n = 0$, $s_i \geqslant 0$, $|s_{i+1} - s_i| = 1$) is

$$-1 + \frac{2}{\sqrt{\pi n}} \sqrt{t(2n - t)} \left\{1 + 0\left(\frac{1}{t}\right) + 0\left(\frac{1}{n - t}\right)\right\}, \quad t \to \infty, \ n - t \to \infty.$$

In [10, 11] Kemp has computed the average height of a prefix of the Dyck-language over a two-letter-alphabet, which can be seen as a path, where the restriction $s_n = 0$ is dropped. (This will sometimes be formulated by "a path leading from $(0, 0)$ to $(n, \cdot)$".) In section 6 the following result is given: The average value s_t of a path leading from $(0, 0)$ to $(n, \cdot)$ is

$$-1 + \sqrt{\frac{8t}{\pi}} \left\{1 + 0\left(\frac{1}{t}\right) + 0\left(\frac{t^3}{n}\right)\right\}, \quad t \to \infty, \ t^3/n \to 0.$$

To obtain s_t Kemp [8] has derived two combinatorial identities. In section 7, first a combinatorial identity is given which generalises an identity in Riordan's book [13] by introducing a parameter. From this identity a further identity is easily derived which contains Kemp's identities as special cases.

A crucial point to obtain all these identities is a certain umbral identity. Some identities are stated which can be derived from this umbral identity.

In section 8 some concluding remarks are made.

2. Enumeration Results Concerning Paths Built up by $-1, 0, 1$

Let $k \geqslant 0$ be fixed and $\varphi_j(y)$ $(0 \leqslant j \leqslant k)$ be the generating function where the coefficient of y^n is the number of paths with height $\leqslant k$ leading to the point (n, j). The recurrences for these functions can be formulated as follows:

$$\begin{pmatrix} 1-y & -y & & & & \\ -y & 1-y & -y & & & \\ & -y & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & -y \\ & & & & -y & 1-y \end{pmatrix} \begin{pmatrix} \varphi_0 \\ \varphi_1 \\ \vdots \\ \varphi_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Now let $a_k(y)$ be the determinant of this matrix with $k + 1$ rows. Expanding, the determinant yields

$$a_{k+2} - (1 - y)a_{k+1} + y^2 a_k = 0, \; a_0 = 1 - y, \; a_1 = 1 - 2y.$$

The solution is

$$a_k(y) = \frac{1}{\sqrt{1-2y-3y^2}}\left[\left(\frac{1-y+\sqrt{1-2y-3y^2}}{2}\right)^{k+2} - \left(\frac{1-y-\sqrt{1-2y-3y^2}}{2}\right)^{k+2}\right]$$

By Cramer's rule

$$\varphi_0(y) = \frac{a_{k-1}(y)}{a_k(y)}.$$

Now $A_k(y) = \sum_{n\geq 0} A_{nk}y^n$ will be written for $\varphi_0(y)$. By the substitution $y = v/(1 + v + v^2)$

$$A_k(y) = \frac{(1/(1+v+v^2))^{k+1} - (v^2/(1+v+v^2))^{k+1}}{(1/(1+v+v^2))^{k+2} - (v^2/(1+v+v^2))^{k+2}}$$
$$= (1 + v + v^2)\cdot\frac{1 - v^{2k+2}}{1 - v^{2k+4}}.$$

Hence

$$A_{nk} = \frac{1}{2\pi i}\int^{(0+)} \frac{dy}{y^{n+1}}(1 + v + v^2)\frac{1 - v^{2k+2}}{1 - v^{2k+4}}$$
$$= \frac{1}{2\pi i}\int^{(0+)} \frac{dv}{v^{n+1}}(1 + v + v^2)^n(1 - v^2)\frac{1 - v^{2k+2}}{1 - v^{2k+4}}.$$

$$B_{nk} := A_{nn} - A_{nk}$$
$$= \frac{1}{2\pi i}\int^{(0+)} \frac{dv}{v^{n+1}}(1 + v + v^2)^n\frac{(1 - v^2)^2 v^{2k+2}}{1 - v^{2k+4}}.$$

The coefficient of v^i in $(1 + v + v^2)^n$ is called trinomial coefficient [3, 12] and is denoted by $\binom{n, 3}{i}$. With this notation

Thus

$$\lambda_1 = \lambda_3 = \lambda_5 = 0,\ \lambda_2 = 1,$$

$$\lambda_4 = (\tfrac{3}{2})^2\kappa_4 = (\tfrac{3}{2})^2(\alpha_4 - 3\alpha_2^2) = \tfrac{3}{2}(1 - 3\cdot\tfrac{2}{3}) = -\tfrac{3}{2},$$

$$\lambda_6 = (\tfrac{3}{2})^3\kappa_6 = (\tfrac{3}{2})^3(\alpha_6 - 15\alpha_2\alpha_4 + 30\alpha_2^3) = (\tfrac{3}{2})^2(1 - 15\cdot\tfrac{2}{3} + 30\,(\tfrac{2}{3})^2) = \tfrac{39}{4}.$$

$\varphi^{(r)}(x) = \dfrac{(-1)^r}{\sqrt{2\pi}} H_r(x)e^{-x^2/2}$, with the Chebyshev-Hermite polynomials of p. 193.

$$3^{-n}\binom{n,\,3}{n+k} = \sqrt{\frac{3}{2n}}\cdot\left\{\varphi + \frac{1}{n}\left(-\frac{3}{2}\right)\frac{1}{24}H_4\cdot\varphi + \frac{1}{n^2}\left(\frac{39}{4.720}H_6\cdot\varphi + \frac{9}{4.1152}H_8\cdot\varphi\right)\right\} + o(n^{-5/2})$$

$\left(\text{The arguments } k\sqrt{\dfrac{3}{2n}} \text{ are omitted.}\right)$

$$= \sqrt{\frac{3}{2n}}\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}k^2\frac{3}{2n}\right)\left\{1 - \frac{1}{16n}\left(k^4\frac{9}{4n^2} - 6k^2\frac{3}{2n} + 3\right) + \frac{13}{960n^2}\left(k^6\frac{27}{8n^3} - 15k^4\frac{9}{4n^2} + 45k^2\frac{3}{2n} + 15\right) + \frac{1}{512n^2}\left(k^8\frac{81}{16n^4} - 28k^6\frac{27}{8n^3} + 210k^4\frac{9}{4n^2} - 420k^2\frac{3}{2n} + 105\right\} + o(n^{-5/2})$$

$$= \sqrt{\frac{3}{4\pi n}}\exp\left(-\frac{3k^2}{4n}\right)\left\{1 - \frac{3}{16n} + \frac{1}{512n^2} + k^2\left(\frac{9}{16n^2} - \frac{81}{256n^3}\right) - k^4\left(\frac{9}{64n^3} - \frac{477}{1024n^4}\right) - k^6\frac{711}{5120n^5} + k^8\frac{81}{8192n^6}\right\} + o(n^{-5/2}).$$

The "o" will be replaced by an "O"; for $k \leqslant n^{1/2}$ the O-term can be essentially distributed in the sum in parentheses.

$$= \sqrt{\frac{3}{4\pi n}}\exp\left(-\frac{3k^2}{4n}\right)\left\{1 - \frac{3}{16n} + \frac{9}{16n^2}k^2 - \frac{9}{64n^3}k^4 + O(n^{-2})\right\}.$$

$$3^{-n}\binom{n,\,3}{n-2k} = \sqrt{\frac{3}{4\pi n}}\exp\left(-\frac{3k^2}{n}\right)\left\{1 - \frac{3}{16n} + \frac{9}{4n^2}k^2 - \frac{9}{4n^3}k^4 + O(n^{-2})\right\}.$$

Now k will be substituted by $k - a$, fixed a. After some manipulations

$$3^{-n}\binom{n,\,3}{n+2(a-k)} = \sqrt{\frac{3}{4\pi n}}\exp\left(-\frac{3k^2}{n}\right)\left\{1 - \frac{3}{16n} - \frac{3a^2}{n} + k\left(\frac{6a}{n} - \frac{45a}{8n^2} - \frac{18a^3}{n^2}\right) + k^2\left(\frac{9}{4n^2} + \frac{18a^2}{n^2}\right) + k^3\left(\frac{45a}{2n^3} + \frac{36a^3}{n^3}\right) - k^4\frac{9}{4n^3} - k^5\frac{27a}{2n^4} + O(n^{-2})\right\}.$$

$$3^{-n}\left[\binom{n,3}{n+2(1-k)}-2\binom{n,3}{n-2k}+\binom{n,3}{n+2(-1-k)}\right]$$
$$=\sqrt{\frac{3}{4\pi n}}\exp\left(-\frac{3k^2}{n}\right)\cdot\left\{-\frac{6}{n}+\frac{36k^2}{n^2}+O(n^{-2})\right\}.$$

For $k > n^{1/2+\epsilon}$ the cited theorem in the general form gives

$$3^{-n}\binom{n,3}{n+k}=O(n^{-m})\quad\text{for all } m\geqslant 1,$$

since in this case the O-term is dominant.

4. Asymptotic Evaluation

First it will be remarked that

$$\zeta^2(z)=\sum_{k\geqslant 1} d(k)/k^z,$$

where $\zeta(z)$ is Riemann's ζ-function. Let

$$h_b(n)=\sum_{k\geqslant 1} k^b d(k)e^{-3k^2/n},\quad \text{fixed b.}$$

As in (2),

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

$$h_b(n)=\sum_{k\geqslant 1}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}k^b d(k)\Gamma(z)\left(\frac{3k^2}{n}\right)^{-z}dz$$
$$=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\left(\frac{n}{3}\right)^z\Gamma(z)\zeta^2(2z-b)\,dz.$$

Thus

$$h_b(n)=g_b\left(\frac{n}{3}\right),$$

where g_b is defined in [2]. Hence for all $m\geqslant 1$

$$h_0(n)=\frac{1}{4}\sqrt{\frac{\pi n}{3}}\left(\log\frac{n}{3}+3\gamma-2\log 2\right)+\frac{1}{4}+O(n^{-m}),$$

$$h_2(n)=\frac{1}{24}n\sqrt{\frac{\pi n}{3}}\left(\log\frac{n}{3}+2+3\gamma-2\log 2\right)+O(n^{-m}).$$

Now consider

$$3^{-n}\sum_{k=1}^{n}d(k)\left[\binom{n,3}{n+2(1-k)}-2\binom{n,3}{n-2k}+\binom{n,3}{n+2(-1-k)}\right].$$

The sum will be splitted into three parts, according to $1\leqslant k\leqslant n^{1/2}$, $n^{1/2}<k\leqslant n^{1/2+\epsilon}$, $n^{1/2+\epsilon}<k\leqslant n$. The third sum is $n^2O(n^{-m})$ for all $m\geqslant 1$. To evaluate the second sum the approximation by the normal distri-

distribution will be used, but the O-term (outside the parentheses) must be regarded.

$$\sum_{n^{1/2}<k\leqslant n^{1/2+\epsilon}} d(k) \leqslant \sum_{1<k\leqslant n^{1/2+\epsilon}} d(k)$$

By an elementary result (due to Dirichlet, see [1]) the last expression is $O(n^{1/2+\epsilon}\log n)$. The first sum gives the main contribution, and here the approximation by the normal distribution will be used, and it makes only an exponently small error to extend the range of summation to infinity. Thus one can continue

$$= \sqrt{\frac{3}{4\pi n}}\Bigg(-\frac{3}{2n}\sqrt{\frac{\pi n}{3}}\left(\log\frac{n}{3}+3\gamma-2\log 2\right)-\frac{3}{2n}$$
$$+\frac{3}{2n}\sqrt{\frac{\pi n}{3}}\left(\log\frac{n}{3}+2+3\gamma-2\log 2\right)+O(n^{-2}h_0(n))\Bigg)$$
$$= \sqrt{\frac{3}{4\pi n}}\left(\frac{3}{n}\sqrt{\frac{\pi n}{3}}-\frac{3}{2n}+O\left(\frac{\log n}{n^{3/2}}\right)\right).$$

Now the number A_{nn} of all nonnegative paths is, as an application of the mirror principle of André (see [4])

$$\binom{n,\,3}{n}-\binom{n,\,3}{n+2} = 3^n\sqrt{\frac{3}{4\pi n}}\left(\frac{3}{n}+O(n^{-3/2}\log n\right).$$

Therefore the desired quantity is

$$\frac{S_n}{A_{nn}} = \left\{\sqrt{\frac{3}{4\pi n}}\cdot\frac{3}{n}\right\}^{-1}\sqrt{\frac{3}{4\pi n}}\left(\frac{3}{n}\sqrt{\frac{\pi n}{3}}-\frac{3}{2n}+O(n^{-3/2}\log n)\right)$$
$$= \sqrt{\frac{\pi n}{3}}-\tfrac{1}{2}+O(n^{-1/2}\log n).$$

The contribution to the sum where $n^{1/2}<k\leqslant n^{1/2+\epsilon}$ is

$$n^{3/2}\cdot O(n^{1/2+\epsilon}\log n)\cdot O(n^{-5/2}) = O(n^{-1/2+\epsilon}\log n),$$

and this is the reason why the ϵ appears in the following

THEOREM 4.1. *The average height of a path from* (0, 0) *to* $(n, 0)$, *built up by* 3 *symbols is*

$$\sqrt{\frac{\pi n}{3}}-\frac{3}{2}+O\left(\frac{\log n}{n^{1/2-\epsilon}}\right),\ n\to\infty, \quad \text{for all } \epsilon>0.$$

5. The Average Height of a Path without Restriction

Let $k\geqslant 0$ be fixed and $\varphi_j(y)$ ($|j|\leqslant k$) be the generating function where the coefficient of y^n is the number of paths with height $\leqslant k$ leading to the point (n, j). The recurrences can be formulated again as a system of linear equations:

$$\begin{pmatrix} 1 & -y & & & & \\ -y & 1 & -y & & & \\ & -y & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & \cdot & 1 & -y \\ & & & & -y & 1 \end{pmatrix} \begin{pmatrix} \varphi_k \\ \vdots \\ \varphi_0 \\ \vdots \\ \varphi_{-k} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

Now let $a_m(y)$ be the determinant of this matrix with m rows. Expanding this determinant yields

$$a_{m+2} - a_{m+1} + y^2 a_m = 0.$$

The solution is

$$a_m(y) = \frac{1}{\sqrt{1-4y^2}}\left[\left(\frac{1+\sqrt{1-4y^2}}{2}\right)^{m+1} - \left(\frac{1-\sqrt{1-4y^2}}{2}\right)^{m+1}\right];$$

$a_m(y^2)$ is a Fibonacci polynomial [6].

By Cramer's rule

$$\varphi_0(y) = \frac{a_k(y)a_k(y)}{a_{2k+1}(y)}.$$

This can be written as

$$\varphi_0(y) = \frac{a_k(y)}{b_k(y)}$$

with

$$b_m(y) = \left(\frac{1+\sqrt{1-4y^2}}{2}\right)^{m+1} + \left(\frac{1-\sqrt{1-4y^2}}{2}\right)^{m+1};$$

$b_m(y^2)$ is a Lucas polynomial:

$$b_m(z) = \sum_{k\geqslant 0} \frac{m}{m-k}\binom{m-k}{k}(-z)^k.$$

Now $A_k(y) = \sum_{n\geqslant 0} A_{nk} y^n$ will be written for $\varphi_0(y)$. By the substitution $y = v/(1+v^2)$

$$A_k(y) = \frac{1+v^2}{1-v^2}\cdot\frac{1-v^{2k+2}}{1+v^{2k+2}}.$$

By the trigonometric change of variable $y = 1/2\cos\theta$

$$A_k(y) = \frac{\cos\theta\cdot\sin(k+1)\theta}{\sin\theta\cdot\cos(k+1)\theta}.$$

An alternative representation for $A_k(y)$ is obtained by partial fractions (as in [6])

$$A_k(y) = \frac{1}{a_k\cdot(k+1)} \sum \frac{1}{1 - y2\cos\dfrac{(2l-1)\pi}{2(k+1)}} + b_k$$

where the sum runs through $l \in \left\{1, \ldots, \left\lfloor \frac{k+1}{2} \right\rfloor, k+2, \ldots, k+1+\left\lfloor \frac{k+1}{2} \right\rfloor\right\}$, with

$$a_k = 2, \qquad b_k = \frac{1}{k+1} \qquad k \text{ odd}$$

$$a_k = k+1,\ b_k = 0 \qquad k \text{ even}.$$

This gives the explicit formula

$$A_{2n,k} = \frac{2}{a_k(k+1)} \sum_{l=0}^{\lfloor (k+1)/2\rfloor - 1} \left(2\cos\frac{(2l+1)\pi}{2(k+1)}\right)^{2n}.$$

By the residue theorem

$$A_{nk} = \frac{1}{2\pi i}\int^{(0+)} \frac{dv}{v^{n+1}}\,(1+v^2)^n\,\frac{1-v^{2k+2}}{1+v^{2k+2}},$$

or, with $v^2 = u$,

$$B_{2n,k} := A_{2n,2n} - A_{2n,k} = \frac{1}{2\pi i}\int^{(0+)} \frac{du}{u^{n+1}}\,(1+u)^{2n}\,\frac{2u^{k+1}}{1+u^{k+1}}.$$

The desired average height is

$$\binom{2n}{n}^{-1} \sum_{k\geq 0} B_{2n,k} = \binom{2n}{n}^{-1} \sum_{k\geq 0}\sum_{l\geq 1} (-1)^{l+1} 2\binom{2n}{n-(k+1)l}$$

$$= \binom{2n}{n}^{-1} \sum_{k\geq 1} [d_1(k) - d_0(k)]2\binom{2n}{n-k},$$

where $d_1(k)$ ($d_0(k)$) denotes the number of odd (even) divisors of k. Now let

$$h_b(n) = \sum_{k\geq 1} k^b[d_1(k) - d_0(k)]\exp(-k^2/n), \text{ fixed } b.$$

Remark that

$$(1-2^{-z})\zeta^2(z) = \sum_{k\geq 1} \frac{d_1(k)}{k^z} \text{ and } 2^{-z}\zeta^2(z) = \sum_{k\geq 1}\frac{d_0(k)}{k^z}.$$

Hence, as in [2]

$$h_b(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \Gamma(z)\left[n^z - \left(\frac{n}{4}\right)^z 2^{b+1}\right]\zeta^2(2z-b)\,dz,\ c > \frac{b+1}{2}.$$

$h_b(n)$ can be expressed as a sum of the residues in (5.1) and (5.2):

$$\tfrac{1}{2}\Gamma\left(\frac{b+1}{2}\right)n^{(b+1)/2}\log 2 \tag{5.1}$$

and

$$\frac{(-1)^k}{k!\, n^k}(1 - 2^{2k+b+1})\frac{B^2_{2k+b+1}}{(2k+b+1)^2},\ k = 0, 1, \ldots .$$

Here the error is $O(n^{-m})$ for all $m > 0$. This gives

$$h_0(n) = \tfrac{1}{2}\sqrt{\pi n}\log 2 - \tfrac{1}{4} + O(n^{-m})$$

$$h_2(n) = \tfrac{1}{4}n\sqrt{\pi n}\log 2 + O(n^{-m})$$

$$h_4(n) = \tfrac{3}{8}n^2\sqrt{\pi n}\log 2 + O(n^{-m}).$$

$$\binom{2n}{n}^{-1}\sum_{k\geq 1}[d_1(k) - d_0(k)]2\binom{2n}{n-k}$$

$$= \sum_{k\geq 1}[d_1(k) - d_0(k)]\exp\left(-\frac{k^2}{n}\right)2\cdot\left[1 + \frac{k^2}{2n^2} + \frac{k^4}{6n^3}\right] + 0(n^{-2+\epsilon}h_0(n))$$

$$= 2h_0(n) + \frac{1}{n^2}h_2(n) - \frac{1}{3n^3}h_4(n) + O(n^{-3/2+\epsilon})$$

$$= (\log 2)\cdot\sqrt{\pi n} - \tfrac{1}{2} + \tfrac{1}{8}(\log 2)\cdot\sqrt{\frac{\pi}{n}} + O(n^{-3/2+\epsilon})$$

$$= (\log 2)\cdot\sqrt{\pi n} - \tfrac{1}{2} + O\left(\frac{1}{\sqrt{n}}\right).$$

This proves

THEOREM 5.1. *The average height of a path from* $(0, 0)$ *to* $(2n, 0)$ *without the restriction* $s_t \geq 0$ *is*

$$(\log 2)\cdot\sqrt{\pi n} - \tfrac{1}{2} + O\left(\frac{1}{\sqrt{n}}\right),\ n \to \infty.$$

6. A Certain Average of a Prefix of the Dyck-Language Over a Two Letter Alphabet

As already mentioned in the introduction, the average value s_t of a path from $(0, 0)$ to $(n, \cdot)$ is desired. Some combinatorial preliminaries are to be made:

Let $(x)_k = x(x-1)\ldots(x-k+1)$ and

$$f_s(n) = \sum_{k=0}^{n}(k)_s\binom{2n}{n-k},\ g_s(n)\sum_{k=0}^{n}(k)_s\binom{2n+1}{n-k}.$$

LEMMA 6.1.

$$f_0(n) = 2^{2n-1} + \binom{2n-1}{n} \qquad g_0(n) = 2^{2n}$$

$$f_1(n) = n\binom{2n-1}{n} \qquad g_1(n) = (2n+1)\binom{2n-1}{n} - 2^{2n-1}$$

$$f_2(n) = n2^{2n-2} - n\binom{2n-1}{n} \qquad g_2(n) = (n+2)2^{2n-1} - (n+1)\binom{2n+1}{n}$$

$$f_3(n) = -3n2^{2n-2} + n(n+2)\binom{2n-1}{n} \qquad g_3(n) = -3(3n+4)2^{2n-2} + (n+6)(2n+1)\binom{2n-1}{n}.$$

Proof. For $0 \leqslant s \leqslant 2$ these identities appear in [13; p. 34]; for $s = 3$ they can be derived in a very similar way.

LEMMA 6.2. (a) $\sum_{k\geqslant 0} (2k)^3\binom{2n}{n-k} = 4n^2\binom{2n}{n}$.

(b) $\sum_{k\geqslant 0} (2k+1)^3\binom{2n+1}{n-k} = (n+1)(4n+1)\binom{2n+1}{n}$.

Proof. (a) This can be derived from Lemma 6.1 if one remarks that

$$(2k)^3 = 8((k)_3 + 3(k)_2 + (k)_1).$$

(b) Remark that

$$(2k+1)^3 = 8(k)_3 + 36(k)_2 + 26(k)_1 + 1.$$

Let $H(n, k, t)$ be the number of paths of length n with $s_t = k$. Then (cf. [8, 9])

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

Now let $Q_t(n)$ be the number of paths from $(0, t)$ to $(n, \cdot)$ (with $s_i \geqslant 0$). Let (as in [12; p. 532]) $g_{nm}^{(t)}$ be the number of paths from $(0, t)$ to (n, m) and

$$G_t(x, z) = \sum_{n\geqslant 0}\sum_{m\geqslant 0} g_{nm}^{(t)} x^m z^n.$$

Since

$$g_{n+1,m}^{(t)} = g_{n,m-1}^{(t)} + g_{n,m+1}^{(t)}, \ m \geqslant 1, n \geqslant 0,$$

$$g_{n+1,0}^{(t)} = g_{n,1}^{(t)}, \qquad g_{0,m}^{(t)} = \delta_{tm},$$

$$G_t(x, z)\cdot\left(1 - \frac{z}{x} - xz\right) = x^t - \frac{z}{x}\cdot g_t(z)$$

with

$$g_t(z) = \sum_{n\geqslant 0} g_{n0}^{(t)} z^n$$

holds. From [13]

$$g_t(z) = \frac{1}{z}\left(\frac{1-\sqrt{1-4z^2}}{2z}\right)^{t+1}.$$

Hence

$$G_t(x, z) = \frac{zg_t(z) - x^{t+1}}{x^2z + z - x} \text{ and } G_t(1, z) = \frac{1 - zg_t(z)}{1-2z}.$$

The usual substitution $z = v/(1+v^2)$ gives

$$Q_t(n) = \frac{1}{2\pi i}\int^{(0+)} \frac{dv}{v^{n+1}}(1+v^2)^n(1+v)(1+v+\ldots+v^t).$$

This proves

LEMMA 6.6

$$Q_t(2n) = \sum_{i=0}^{\lfloor t/2 \rfloor} \binom{2n}{n-i} + \sum_{i=1}^{\lfloor (t+1)/2 \rfloor} \binom{2n}{n-i},$$

$$Q_t(2n+1) = \sum_{i=0}^{\lfloor (t+1)/2 \rfloor} \binom{2n+1}{n-i} + \sum_{i=0}^{\lfloor t/2 \rfloor} \binom{2n+1}{n-i}.$$

This lemma gives the following asymptotic equivalent for the number of paths from (0, 0) to $(n, \cdot)$:

$$\binom{n}{\lfloor n/2 \rfloor}(1+k)\left(1+O\left(\frac{k^2}{n}\right)\right) = \binom{n}{\lfloor n/2 \rfloor}\left(1+k+O\left(\frac{k^3}{n}\right)\right) \tag{6.1}$$

provided that $n \to \infty$, $k^3/n \to 0$. (This can be derived by means of Stirling's formula (see [2]).

Hence

$$H(n-1, k-1, t-1) = \frac{k}{t}\binom{t}{(t+k)/2}\binom{n-t}{\lfloor (n-t)/2 \rfloor}\{k + 0(k^3/n)\}.$$

Remark that

$$H(n, 0, 0) = \binom{n}{\lfloor n/2 \rfloor}.$$

The desired expected value $R(n-1, t-1)$ is

$$\begin{aligned}
R(n-1, t-1) &= (H(n-1, 0, 0))^{-1} \sum_{k \geq 1} (k-1)H(n-1, k-1, t-1) \\
&= -1 + (H(n-1, 0, 0))^{-1} \sum_{k \geq 1} kH(n-1, k-1, t-1) \\
&= -1 + \binom{n-1}{\lfloor (n-1)/2 \rfloor}^{-1} \sum_{k \geq 1} \frac{k^3}{t}\binom{t}{(t+k)/2}\binom{n-t}{\lfloor (n-t)/2 \rfloor} \\
&\quad \cdot \{1 + O(k^2/n)\} \\
&= -1 + \binom{n-1}{\lfloor (n-1)/2 \rfloor}^{-1} \binom{n-t}{\lfloor (n-t)/2 \rfloor} \frac{1}{t} \\
&\quad \cdot \sum_{k \geq 1} k^3 \binom{t}{(t+k)/2}\{1 + O(k^2/n)\}.
\end{aligned}$$

Now the sum will be evaluated. First, let $t = 2T$; clearly k must be even and the sum becomes

$$\sum_{k \geq 1} (2k)^3 \binom{2T}{T-k} = 4T^2\binom{2T}{T}.$$

Now let $t = 2T + 1$; then the sum becomes

$$\sum_{k \geq 0} (2k+1)^3 \binom{2T+1}{T-k} = (T+1)(4T+1)\binom{2T+1}{T}.$$

The lemmata 6.1 and 6.2 suggest that

$$\sum_{k \geq 1} k^b \binom{t}{(t+k)/2} = O\left(t^{(b+1)/2} \cdot \binom{t}{\lfloor t/2 \rfloor}\right)$$

holds, and this can be settled for $b = 5$ by a direct computation. Hence

$$R(n-1, t-1) = -1 + \binom{n-1}{\lfloor (n-1)/2 \rfloor}^{-1} \binom{n-t}{\lfloor (n-t)/2 \rfloor} \frac{1}{t} t^2 \binom{t}{\lfloor t/2 \rfloor} \cdot \{1 + 0(t^3/n)\}.$$

Now assume $t \to \infty$, $t^3/n \to O$.

$$R(n-1, t-1) = -1 + \sqrt{\pi \frac{n-1}{2}}\, 2^{-(n-1)} \sqrt{\frac{2}{\pi(n-t)}}\, 2^{n-t} t \sqrt{\frac{2}{\pi t}}\, 2^t \cdot \left\{1 + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n-t}\right) + O\left(\frac{1}{t}\right) + O\left(\frac{t^3}{n}\right)\right\}.$$

$$= -1 + \sqrt{\frac{2}{\pi}} \sqrt{\frac{n-1}{t(n-t)}}\, 2t \left\{1 + O\left(\frac{1}{t}\right) + O\left(\frac{t^3}{n}\right)\right\}.$$

$$R(n, t) = -1 + \sqrt{\frac{8}{\pi}} \sqrt{\frac{tn}{n-t}} \left\{1 + O\left(\frac{1}{t}\right) + O\left(\frac{t^3}{n}\right)\right\}$$

$$= -1 + \sqrt{\frac{8t}{\pi}} \left\{1 + O\left(\frac{1}{t}\right) + O(t^3/n)\right\}.$$

7. Combinatorial Identities which are Related to the Considered Problems

First, an identity will be proved, which is a generalization (for $s = 0$) of an identity in Riordan's book [13; p. 89].

THEOREM 7.1

$$\sum_{k \geq 1} (2k + s) \binom{2m+s}{m-k} \binom{2n+s}{n-k} = \binom{2m+s}{m} \binom{2n+s}{n} \frac{mn}{m+n+s}.$$

Proof. Let $s \in \mathbb{N}_0$ and $d_n(s, m)$ be a sequence defined by

$$d_0(s, m) = m^{-1}$$

$$d_n(s, m) = \sum_{k \geq 0} \frac{2n+s}{n+k+s} \binom{n+k+s}{2k+s} \binom{2k+s}{k+s} \frac{(-1)^k}{k+m} \quad \text{for } n \geq 1.$$

Kemp [8] has shown that the inverse relation is

$$\frac{1}{n+m} \binom{2n+s}{n+s} = \sum_{k \geq 0} (-1)^k \binom{2n+s}{n-k} d_k(s, m).$$

Furthermore he has computed the explicit form

$$d_n(s, m) = \begin{cases} m^{-1} & \text{for } n = 0 \\ (-1)^{n-1}(2n+s) \dfrac{(m-s-1)!\,(m-1)!}{(m-s-n)!\,(m+n)!} & \text{for } n \geq 1. \end{cases}$$

Substituting this form in the inverse relation yields

$$\frac{1}{n+m}\binom{2n+s}{n+s} = \frac{1}{m}\binom{2n+s}{n} - \sum_{k\geq 1}\binom{2n+s}{n-k}(2k+s)\frac{(m-s-1)!\,(m-1)!}{(m-s-k)!\,(m+k)!}.$$

Replacing m by $m+s$ this yields

$$\binom{2n+s}{n}\frac{n}{(m+s)(n+m+s)} = \sum_{k\geq 1}(2k+s)\binom{2n+s}{n-k}\binom{2m+s}{m-k}\frac{1}{m(m+s)}\binom{2m+s}{m}^{-1}.$$

This expression is equivalent to the proposition.

This identity is useful to derive an identity, which contains two identities of Kemp [8] as special cases ($s = 0, 1$). Kemp has used them to derive the results concerning the average value of s_t (see section 6).

THEOREM 7.2

$$\sum_{k\geq 1}(2k+s)^3\binom{2m+s}{m-k}\binom{2n+s}{n-k} = \binom{2m+s}{m}\binom{2n+s}{n}\frac{mn\{s^2(m+n+s+3)+4(mn+ms+ns)\}}{(m+n+s)(m+n+s-1)}.$$

Proof. Let $f_s(m, n)$ be the expression in theorem 7.1 and $g_s(m, n)$ the expression in theorem 7.2. Since

$$\binom{2(m-1)+s}{m-1-k} = \frac{(m-k)(m+k+s)}{(2m-1+s)(2m+s)}\binom{2m+s}{m-k}$$

and

$$4(m-k)(m+k+s) = (2m+s)^2 - (2k+s)^2,$$

the equation

$$4(2m-1+s)(2m+s)f_s(m-1, n) = (2m+s)^2 f_s(m, n) - g_s(m, n)$$

is obtained. Hence

$$\begin{aligned} g_s(m, n) &= (2m+s)^2\binom{2m+s}{m}\binom{2n+s}{n}\frac{mn}{m+n+s} - 4(2m-1+s) \\ &\quad \times (2m+s)\binom{2(m-1)+s}{m-1}\binom{2n+s}{n}\frac{(m-1)n}{m-1+n+s} \\ &= \binom{2m+s}{m}\binom{2n+s}{n}\left\{(2m+s)^2\frac{mn}{m+n+s} - 4m(m+s)\frac{(m-1)n}{m-1+n+s}\right\} \\ &= \binom{2m+s}{m}\binom{2n+s}{n}\frac{mn}{(m+n+s)(m+n+s-1)} \\ &\quad \times\{(2m+s)^2(m+n+s-1) - 4(m+s)(m-1)(m+n+s)\}. \end{aligned}$$

Furthermore

$$\begin{aligned}\{\ldots\} &= (4m^2+4ms+s^2)(m+n+s-1)\\ &\quad -(4m^2+4ms-4m-4s)(m+n+s)\\ &= -(4m^2+4ms)+s^2(m+n+s-1)+(4m+4s)(m+n+s)\\ &= s^2(m+n+s-1)+(4m+4s)(n+s)\\ &= s^2(m+n+s+3)+4(mn+ms+ns).\end{aligned}$$

A crucial point in the derivation of these identities is the umbral identity

$$\frac{(n-k)!\,(m+k-1)!}{(m+n)!} = x^k(1-x)^{n-k},\ x^k \equiv x_k(m) = \frac{1}{k+m}. \quad (7.1)$$

which is needed in the computations of Kemp concerning the $d_k(s, m)$.

In the rest of this section some identities are listed which can be easily derived from other identities by this umbral identity.

THEOREM 7.3.

$$\binom{m+n}{m}\sum_{k=0}^{n}(-1)^k\binom{n}{k}^2\frac{m}{k+m} = \sum_{k=0}^{N}(-1)^k\binom{m+n}{k}\binom{m+k-1}{k},\ N=\left\lfloor\frac{n}{2}\right\rfloor.$$

Proof. From

$$\binom{n}{k}^2 = \sum_{j\geqslant 0}\frac{n!}{(k-j)!\,(n-j-k)!\,j!\,j!} \quad [13;\ \text{p. } 41]$$

it follows

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}^2 x^k = \sum_{j=0}^{N}\frac{(-1)^j n!}{j!\,j!\,(n-2j)!}\,x^j(1-x)^{n-2j}. \quad (7.2)$$

A slight modification of the umbral identity (7.1) yields

$$\frac{(n-2k)!\,(m+k-1)!}{(m+n-k)!} = x^k(1-x)^{n-2k},\ x^k \equiv x_k(m) = \frac{1}{k+m}, \quad (7.3)$$

and a substitution of (7.1) in (7.2) and some rearrangements give the proof.

THEOREM 7.4.

$$\binom{m+n}{m}\sum_{k=0}^{n}(-1)^k\binom{n}{k}^3\frac{m}{k+m}$$
$$= \sum_{k=0}^{N}(-1)^k\binom{m+n}{k}\binom{m+k-1}{k}\binom{n+k}{k},\ N=\left\lfloor\frac{n}{2}\right\rfloor.$$

Proof. Starting from

$$\binom{n}{k}^3 = \sum_{j\geqslant 0}\frac{(n+j)!}{(n-k-j)!\,(k-j)!\,j!\,j!\,j!} \quad [13;\ \text{p. } 41]$$

one obtains

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k}^3 x^k = \sum_{j=0}^{N} (-1)^j \frac{(n+j)!}{j!\, j!\, j!\, (n-2j)!} x^j (1-x)^{n-2j}, \quad (7.4)$$

and the result follows again by the substitution of (7. 3) in (7.4).

If one starts not with $\binom{n}{k}^2$ or $\binom{n}{k}^3$ but with $\binom{n}{k}$ one gets

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} x^k = (1-x)^n,$$

and thus

$$\binom{m+n}{m} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{m}{k+m} = 1,$$

but this identity is well-known [13; p. 29].

THEOREM 7.5.

$$\binom{m+n}{m} \sum_{k=0}^{n} (-1)^k \binom{n+k}{2k} \binom{2k}{k} \frac{m}{k+m} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{m+k-1}{k}.$$

Proof. From [13; p. 78]

$$\sum_{k=0}^{n} \binom{n+k}{2k} \binom{2k}{k} \cdot x^k = \sum_{k=0}^{n} \binom{n}{k}^2 x^k (1+x)^{n-k}.$$

Replacing x by $-x$ an application of (7.1) gives the result.

THEOREM 7.6.

$$\binom{m+n}{m} \sum_{k=0}^{n} \binom{n}{k}^2 \frac{m}{k+m} = \sum_{k=0}^{n} \binom{n+k}{k} \binom{m+k-1}{k}.$$

Proof. From [13; p. 81]

$$\sum_{k=0}^{n} \binom{n}{k}^2 x^k = \sum_{k=0}^{n} \binom{n+k}{2k} \binom{2k}{k} x^k (1-x)^{n-k}.$$

An application of (7.1) yields the result.

THEOREM 7.7.

$$\sum_{k=0}^{n} \binom{n+p}{k+p} (-1)^k \frac{m}{m+k} = \sum_{k=0}^{n} \binom{p+k-1}{k} \cdot \binom{m+n-k}{m}^{-1}.$$

Proof. A replacement in (7.1) of $k := 0$ and $n := n-k$ yields

$$\frac{(n-k)!\,(m-1)!}{(m+n-k)!} = (1-x)^{n-k},\ x^k \equiv x_k(m) = \frac{1}{k+m}.$$

The identity [13; p. 47]

$$\sum_{k=0}^{n} \binom{n+p}{k+p} (-1)^k x^k = \sum_{k=0}^{n} \binom{p+k-1}{k} (1-x)^{n-k}$$

leads to the result.

It seems to be evident that there are other identities that can be obtained in this way.

8. Concluding Remarks

The method of obtaining the generating functions in sections 2, 4 by means of a system of linear equations yields to the possibility to make all computations more generally, namely for paths from $(0, t_1)$ to (n, t_2). Furthermore it is possible to introduce probabilities, that means that $s_{l+1} - s_l = 1$ with probability p and $= -1$ with probability $q = 1 - p$ and similar in the case of 3 directions.

In section 6 the assumption $t^3/n \to 0$ is made. If this assumption is dropped, the asymptotic equivalent (6.1) changes and the result will be another one, depending on other assumptions. But the corresponding computations seem to be not essentially different from the computations in this paper.

In section 7 the identity with exponent 3 is obtained from that one with exponent 1. The same trick can be used to obtain the identities with exponents 5, 7, 9, ...

The author wants to thank D. E. Knuth for the announcing of reference [7] and W. Panny for finding out several errors in a previous version of this paper.

References

[1] T. M. Apostol (1976), *Introduction to Analytic Number Theory*, Springer, New York.

[2] N. G. de Bruijn, D. E. Knuth and S. O. Rice (1972), "The average height of planted plane trees", in *Graph Theory and Computing* (R. C. Read, ed.), Academic Press, New York, 15–22.

[3] L. Comtet (1974), *Advanced Combinatorics*, Reidel Publishing Company, Dordrecht, Boston.

[4] W. Feller (1957), *An Introduction to Probability Theory and its Applications*, Wiley, New York.

[5] P. Flajolet (1978), "Analyse d'algorithmes de manipulation de fichiers", IRIA-laboria, Rapport de Recherche No. 321.

[6] P. Flajolet, J. C. Raoult and J. Vuillemin (1979), "The number of registers required for evaluating arithmetic expressions", *Theoretical Computer Science*, **9**, 99–126.

[7] B. W. Gnedenko and A. N. Kolmogorov (1959), *Grenzverteilungen von Summen unabhängiger Zufallsgrößen*, *Akademie*-Verlag, Berlin.

[8] R. Kemp (1979), "On the average stack size of regularly distributed binary trees", in Proceedings of ICALP 79 (Graz, H. Maurer, ed.), *Lecture Notes in Computer Science*, Vol. 71, Springer, New York.

[9] R. Kemp (1979), "A note on the stack size of regularly distributed binary trees", BIT, **20**, 157–163.

[10] R. Kemp (1979), "The average depth of a prefix of the Dycklanguage D_1", in Proceedings of the 2nd Int. Conf. on Fundamentals of Computation Theory (FCT) (L. Budach, ed.), Akademie-Verlag, Berlin.

[11] R. Kemp (1979), "On the average depth of a prefix of the Dycklanguage D_1", Preprint, Univ. Saarbrücken.

[12] D. E. Knuth (1973), *The Art of Computer Programming*, Vol. 1, Addison-Wesley, Reading.

[13] J. Riordan (1968), *Combinatorial Identities*, Wiley, New York.

[Received, May 20, 1980]