Studia Scientiarum Mathematicarum Hungarica 20 (1985), 119—132

THE EXPECTED HEIGHT OF PATHS FOR SEVERAL NOTIONS OF HEIGHT

WOLFGANG PANNY and HELMUT PRODINGER

Abstract

In this paper lattice paths with two directions are considered. Several notions of height are introduced, namely the maximal deviation, the maximal span and the onesided height. Assuming all paths of length n to be equally likely, exact enumeration formulae for the expected height and their asymptotic equivalents are derived.

1. Introduction

This paper deals with the *expected height of lattice paths* for several notions of height.

A *path of length* n is a sequence of integers $a_0, a_1, \dots, a_n$ with $|a_{i+1}-a_i|=1$, $0 \leqq i < n$.

Let us just review some previously known results:

If all paths $a_0, a_1, \dots, a_{2n}$ with $a_0=a_{2n}=0$, $a_i \geqq 0$ are assumed to be equally likely and the height of the path is defined to be $\max\{a_i \mid 0 \leqq i \leqq 2n\}$, then the expected height is

$$\sqrt{\pi n} - \frac{3}{2} + O(n^{-1/2+\varepsilon}) \quad \text{for} \quad \varepsilon > 0 \quad \text{and} \quad n \to \infty. \tag{1.1}$$

This result is due to De Bruijn, Knuth and Rice [2] and was stated not in terms of paths but in terms of *planted plane trees*. Such a path can also be considered as a *Dyckword* of length $2n$, if the i-th letter of the word is an opening (closing) bracket for $a_i - a_{i-1} = 1$ (-1).

A word u is said to be a *prefix* of a word w iff there exists a word v with $uv=w$. So a prefix of a Dyckword can be considered as a path $a_0, a_1, \dots, a_n$ with $a_0=0$ and $a_i \geqq 0$. Assuming all such paths to be equally likely and defining the height again by $\max\{a_i \mid 0 \leqq i \leqq n\}$, the expected height is

$$(\log 2)\sqrt{2\pi n} - \frac{3}{2} + O(n^{-1/2}) \quad \text{for} \quad n \to \infty. \tag{1.2}$$

This result is due to Kemp [6].

Regarding paths where a *third direction* is allowed $(a_{i+1}-a_i=0)$, see [10] and [9].

For a fairly exhaustive list of references of related problems see [7].

1980 *Mathematics Subject Classification*. Primary 05C05; Secondary 68E10.
Key words and phrases. Lattice paths, generating functions, asymptotic expansions.

In Section 2 we consider paths $a_0, \ldots, a_n$ with $a_0=0$. The notion of height is defined by $\max\{|a_i| \mid 0 \leqq i \leqq n\}$ and is called *maximal deviation*. This class of paths corresponds to the prefixes of the Dycklanguage, except for the condition $a_i \geqq 0$, which need not be fulfilled now. We prove that the expected maximal deviation is given by

$$(1.3) \qquad \sqrt{\frac{n\pi}{2}} - \frac{1}{2} + O(n^{-1/2+\varepsilon}) \quad \text{for} \quad \varepsilon > 0 \quad \text{and} \quad n \to \infty.$$

Though the result resembles the previous ones, it cannot be concluded from them as a corollary; to handle the problem we had to use a new technique: Unlike in the former problems, we had not to approximate one single binomial coefficient by the exponential function but a sum of binomial coefficients by the error function. Thus the *Mellin transform* of the error function and the inversion formula of the Mellin transform come into play.

In Section 3 the same family of paths is considered, but the height is now $\max\{a_i - a_j \mid 0 \leqq i, j \leqq n\}$, called *maximal span*. We prove that the expected maximal span is given by

$$(1.4) \qquad \sqrt{\frac{8n}{\pi}} - 1 + O(n^{-1/2}) \quad \text{for} \quad n \to \infty.$$

Though the explicit enumeration formulae are even more complicated than in Section 2, the derivation of the asymptotic formula (1.4) is (due to lucky circumstances) quite elementary.

We would like to mention that the maximal span preserves a way to give a meaningful notion of height not only for Dyckwords of prefixes or Dyckwords but *for all words over a two-letter-alphabet*!

In Section 4 we consider paths $a_0, a_1, \ldots, a_{2n}$ with $a_0 = a_{2n} = 0$ and the *one-sided height* defined by $\max\{a_i \mid 0 \leqq i \leqq 2n\}$. We prove that the expected onesided height is given by

$$(1.5) \qquad \frac{1}{2}\sqrt{n\pi} - \frac{1}{2} + O(n^{-1/2}) \quad \text{for} \quad n \to \infty.$$

Again the derivation of the asymptotic formula is — in a certain sense — elementary. The following interpretation can be given: Suppose there are two players A and B each one having n cards out of the set $\{1, \ldots, 2n\}$. A always leads a card and B follows. A player makes a trick whenever his number is the greater one. The interesting parameter is *the number of tricks that player A can make*, if B plays his optimal strategy. The partitioning of the cards can be seen as a path by defining $a_0=0$ and $a_i - a_{i-1} = 1$ (-1) iff player B (A) has the card i. For example, A has cards 2, 4, 5, 7 means the path 0, 1, 0, 1, 0, —1, 0, —1, 0. It is not hard to see that the number of tricks that player A can make is just the onesided height of the corresponding path!

Section 5 is devoted to some concluding remarks.

In the following sections the *terminus* path is used according to the introduction.

We would like to point out that one notation may have different meanings in different sections.

2. The expected maximal deviation

Let $\psi_{h,l}(z)$ be the generating function where the coefficient of z^n gives the number of paths from (0, 0) to (n, l) with maximal deviation $\leqq h$.

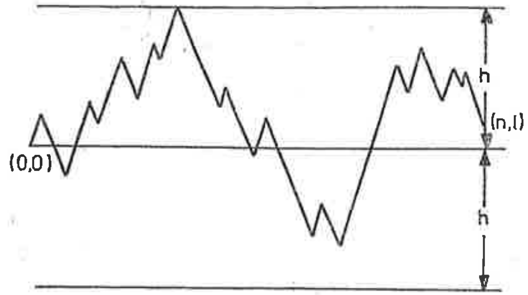

Fig. 1

THEOREM 2.1. *With* $z = v/(1+v^2)$ *we have*

$$(2.1) \qquad \psi_{h,l}(z) = \frac{1+v^2}{1-v^2}\, v^l\, \frac{1-v^{2(h+1-l)}}{1+v^{2h+2}}.$$

PROOF. Regarding the last step of the path we find the following system of linear recurrences for the generating functions $\psi_{h,l}$ $(|l| \leqq h)$ which can be expressed in matrix form:

$$(2.2) \qquad \begin{bmatrix} 1 & -z & & & & & \\ -z & 1 & -z & & & & \\ & \ddots & \ddots & & & & \\ & & -z & 1 & -z & & \\ & & & \ddots & \ddots & & \\ & & & & -z & 1 & -z \\ & & & & & -z & 1 \end{bmatrix} \cdot \begin{bmatrix} \psi_{h,-h}(z) \\ \psi_{h,-h+1}(z) \\ \vdots \\ \psi_{h,0}(z) \\ \vdots \\ \psi_{h,h}(z) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Compare [10], where a similar system is used.

Using Cramer's rule we get

$$(2.3) \qquad \psi_{h,l}(z) = \frac{z^{|l|}\, a_{h-1}(z)\, a_{h-l-1}(z)}{a_{2h}(z)},$$

where $a_i(z)$ denotes the determinant of the matrix in (2.2) with $i+1$ rows. From [6], [9], [10] we know

$$(2.4) \qquad a_i(z) = \frac{1}{1-v^2} \frac{1-v^{2i+4}}{(1+v^2)^{i+1}},$$

where the substitution $z = v/(1+v^2)$ is used. Inserting (2.4) in (2.3) we get the desired result. □

Let $\Psi_h(z)=\sum_{|l|\leq h}\psi_{h,l}(z)$ be the generating function of the number $c_{n,h}$ of all paths with maximal deviation $\leq h$.

THEOREM 2.2.

$$\Psi_h(z)=\frac{(1+v^2)(1-v^{h+1})^2}{(1-v)^2(1+v^{2h+2})}. \tag{2.5}$$

PROOF.

$$\Psi_h(z)=\psi_{h,0}(z)+2\sum_{1\leq l\leq h}\psi_{h,l}(z).$$

Using Theorem 2.1 we get the result by an elementary computation. □

THEOREM 2.3.

$$c_{n,h}=2^n-2\sum_{\lambda\geq 0}\sum_{0\leq l\leq h}\left[\binom{n}{\left\lceil\frac{n-(h+2)}{2}\right\rceil-2\lambda(h+1)-l}+\binom{n}{\left\lfloor\frac{n-(h+2)}{2}\right\rfloor-2\lambda(h+1)-l}\right]. \tag{2.6}$$

PROOF. Using Cauchy's integral formula we have

$$c_{n,h}=\frac{1}{2\pi i}\int^{(0+)}\frac{dz}{z^{n+1}}\Psi_h(z)=\frac{1}{2\pi i}\int^{(0+)}\frac{dv(1+v^2)^n(1+v)(1-v^{h+1})^2}{v^{n+1}(1-v)(1+v^{2h+2})} \tag{2.7}$$

where the substitution $z=v/(1+v^2)$ was used. Hence $c_{n,h}$ is the coefficient of v^n in

$$\frac{(1+v^2)^n(1+v)(1-v^{h+1})^2}{(1-v)(1+v^{2h+2})}. \tag{2.8}$$

Expanding the denominator we get

$$\frac{1}{(1-v)(1+v^{2h+2})}=\sum_{\lambda\geq 0}\sum_{l=0}^{2h+1}v^{4\lambda(h+1)+l}.$$

The result follows now by some elementary manipulations. □

As an alternative representation one obtains in a similar way:

THEOREM 2.4. *With the abbreviation*

$$B(n,k)=\sum_{l=0}^{k}\binom{n}{l},$$

$$c_{n,h}=2\sum_{\lambda\geq 0}(-1)^\lambda\left[B\left(n,\left\lceil\frac{n-(h+2)}{2}\right\rceil-\lambda(h+1)\right)+B\left(n,\left\lfloor\frac{n-(h+2)}{2}\right\rfloor-\lambda(h+1)\right)\right]. \quad □ \tag{2.9}$$

A further representation for $c_{n,h}$ can be obtained from (2.8) by partial fraction expansion.

THEOREM 2.5.

$$c_{n,h}=\frac{2^n}{h+1}\sum_{0\leq l\leq h}(-1)^l\cos^n\left(\frac{2l+1}{2(h+1)}\pi\right)\operatorname{ctg}\left(\frac{2l+1}{4(h+1)}\pi\right). \quad □ \tag{2.10}$$

COROLLARY 2.6. *For $n\leq h$ we have*

$$\sum_{0\leq l\leq h}(-1)^l\cos^n\left(\frac{2l+1}{2(h+1)}\pi\right)\operatorname{ctg}\left(\frac{2l+1}{4(h+1)}\pi\right)=h+1. \quad □ \tag{2.11}$$

Since the total of paths of length n is 2^n, we immediately obtain the following corollary:

COROLLARY 2.7. *With $d_{n,h}$ denoting the number of paths of length n and maximal deviation $>h$, we have*

$$d_{n,h}=2\sum_{\lambda\geq 0}\sum_{0\leq l\leq h}\left[\binom{n}{\left\lceil\frac{n-(h+2)}{2}\right\rceil-2\lambda(h+1)-l}+\binom{n}{\left\lfloor\frac{n-(h+2)}{2}\right\rfloor-2\lambda(h+1)-l}\right]. \quad □ \tag{2.12}$$

Using Abel's summation formula, we get the following exact expression for D_n, the expected maximal deviation of a path of length n.

THEOREM 2.8.

$$D_n=2^{-n}2\sum_{h\geq 0}\sum_{\lambda\geq 0}\sum_{0\leq l\leq h}\left[\binom{n}{\left\lceil\frac{n-(h+2)}{2}\right\rceil-2\lambda(h+1)-l}+\binom{n}{\left\lfloor\frac{n-(h+2)}{2}\right\rfloor-2\lambda(h+1)-l}\right]. \quad □ \tag{2.13}$$

The remainder of this section is devoted to the study of the asymptotic behaviour of D_n. For the sake of brevity we confine ourselves to the case of even n.

As a first step we have to approximate sums of binomial coefficients of the following kind $(0\leq a\leq b)$:

$$S_{a,b}:=2^{-2n}\sum_{a\leq k\leq b}\binom{2n}{n+k}. \tag{2.14}$$

THEOREM 2.9. *Let $\varepsilon>0$. Assume $0\leqq a\leqq b=O(n^{1/2+\varepsilon})$ and $k=O(n^{1/2+\varepsilon})$. Furthermore let* erfc(x) *be the well-known complement of the error function:*

$$\operatorname{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\,dt.$$

Then

$$\text{(i)}\quad S_{a,b}=\frac{1}{2}\left[\operatorname{erfc}\left(\frac{a-\frac{1}{2}}{\sqrt{n}}\right)-\operatorname{erfc}\left(\frac{b+\frac{1}{2}}{\sqrt{n}}\right)\right](1+O(n^{-1+\varepsilon})),$$

$$\text{(2.15)}\quad \text{(ii)}\quad S_{k,k}=\left[\operatorname{erfc}\left(\frac{k}{\sqrt{n}}\right)-\operatorname{erfc}\left(\frac{k+\frac{1}{2}}{\sqrt{n}}\right)+\frac{1}{2}\frac{ke^{-k^2/n}}{\sqrt{\pi n^3}}\right](1+O(n^{-1+\varepsilon})),$$

$$\text{(iii)}\quad S_{k,k}=\left[\operatorname{erfc}\left(\frac{k-\frac{1}{2}}{\sqrt{n}}\right)-\operatorname{erfc}\left(\frac{k}{\sqrt{n}}\right)-\frac{1}{2}\frac{ke^{-k^2/n}}{\sqrt{\pi n^3}}\right](1+O(n^{-1+\varepsilon})).$$

PROOF. For the sake of brevity we only want to stress the main ideas of the proof. Following the presentation in [5, pp 179—182] one can distinguish three components of the error committed in the above approximations. The first component is effected by the approximation

$$2^{-2n}\binom{2n}{n+k}\sim 2^{-2n}\binom{2n}{n}e^{-k^2/n}.$$

The second component is due to the approximation of the middle term by Stirling's formula

$$2^{-2n}\binom{2n}{n}\sim\frac{1}{\sqrt{n\pi}}.$$

For this two error components it follows easily from the estimates given in [5] that we have

$$2^{-2n}\binom{2n}{n+k}=\frac{1}{\sqrt{n\pi}}e^{-k^2/n}(1+O(n^{-1+\varepsilon}))\quad\text{for}\quad k=O(n^{(1/2)+\varepsilon}).$$

The third component is caused by replacing summation over $a\leqq k\leqq b$ by integration within the bounds $a-\frac{1}{2}$ and $b+\frac{1}{2}$. Regarding part (i) of the above theorem it suffices to consider the approximation

$$\frac{1}{\sqrt{2\pi}}\sqrt{\frac{2}{n}}\,e^{-k^2/n}\sim\frac{1}{\sqrt{2\pi}}\int_{(k-\frac{1}{2})\sqrt{\frac{2}{n}}}^{(k+\frac{1}{2})\sqrt{\frac{2}{n}}}e^{-t^2/2}\,dt.$$

Estimating the difference between these two terms by appropriate chosen triangles we are led to (i). As for part (ii) and (iii) a similar reasoning can be used to derive the indicated correction terms and to estimate the order of the error. □

The correction terms in the formulae (ii) and (iii) allow us to lower the order of the error from $O(n^{-1/2+\varepsilon})$ to $O(n^{-1+\varepsilon})$.

THEOREM 2.10. *Let σ and τ be the arithmetical functions defined by*

$$\text{(2.16)}\qquad \sigma(m)=\sum_{\substack{m=(4\lambda+1)(h+1),\\ \lambda,h\geqq 0}}1,\qquad \tau(m)=\sum_{\substack{m=(4\lambda+3)(h+1)\\ \lambda,h\geqq 0}}1.$$

Then

$$\text{(2.17)}\qquad \begin{aligned} D_{2n}&=2\sum_{m\geqq 1}\left[(\sigma(m)-\tau(m))\operatorname{erfc}\left(\frac{m}{2\sqrt{n}}\right)\right](1+O(n^{-1+\varepsilon}))+\\ &+\frac{1}{\sqrt{\pi n^3}}\sum_{m\geqq 1}[(\sigma(m)-\tau(m))me^{-m^2/n}](1+O(n^{-1+\varepsilon})).\end{aligned}$$

PROOF. Starting from (2.13), considering the cases h even or odd separately and regarding that $\binom{2n}{n-k}=\binom{2n}{n+k}$, we have

$$\text{(2.18)}\qquad \begin{aligned} D_{2n}&=2^{-2n}2\sum_{\substack{h,\lambda\geqq 0\\ h\text{ even}}}\sum_{0\leqq l\leqq h}2\binom{2n}{n+\frac{h+2+4\lambda(h+1)+2l}{2}}+\\ &+2^{-2n}2\sum_{\substack{h,\lambda\geqq 0\\ h\text{ odd}}}\left[\sum_{0\leqq l\leqq h-1}2\binom{2n}{n+\frac{h+3+4\lambda(h+1)+2l}{2}}+\right.\\ &\left.+\binom{2n}{n+\frac{h+1+4\lambda(h+1)}{2}}+\binom{2n}{n+\frac{3h+3+4\lambda(h+1)}{2}}\right].\end{aligned}$$

Now we approximate the last two terms in (2.8) by (ii) and (iii) of Theorem 2.9, respectively, and the remaining sums by (i) and obtain:

$$\begin{aligned} D_{2n}&=2\sum_{h,\lambda\geqq 0}\left[\operatorname{erfc}\left(\frac{h+1+4\lambda(h+1)}{2\sqrt{n}}\right)-\operatorname{erfc}\left(\frac{3h+3+4\lambda(h+1)}{2\sqrt{n}}\right)\right](1+O(n^{-1+\varepsilon}))+\\ &+\frac{2}{\sqrt{2\pi}}\sum_{k,\lambda\geqq 0}\left[\frac{k+1+4\lambda(k+1)}{\sqrt{2n^3}}\exp\left(\frac{-[k+1+4\lambda(k+1)]^2}{n}\right)-\right.\\ &\left.-\frac{3k+3+4\lambda(k+1)}{\sqrt{2n^3}}\exp\left(\frac{-[3k+3+4\lambda(k+1)]^2}{n}\right)\right](1+O(n^{-1+\varepsilon})).\end{aligned}$$

The error committed in the above approximation by extending the range of summation to infinity is exponentially small and therefore covered by the error terms.

The desired result now immediately follows by use of the arithmetical functions σ and τ. □

The next lemma deals with the generating Dirichlet series of $\sigma(m)-\tau(m)$:

LEMMA 2.11.

$$\sum_{m\geq 1}\frac{\sigma(m)-\tau(m)}{m^z}=\frac{1}{4^z}\zeta(z)\left[\zeta\left(z,\frac{1}{4}\right)-\zeta\left(z,\frac{3}{4}\right)\right], \tag{2.19}$$

where $\zeta(z)$ is the zeta function of Riemann and $\zeta(z,a)$ is the zeta function of Hurwitz (cf. e.g. [1], [12]).

PROOF. First we compute

$$\sum_{m\geq 1}\frac{\sigma(m)}{m^z}=\sum_{m\geq 1}m^{-z}\sum_{\substack{m=(4\lambda+1)(h+1)\\ \lambda,h\geq 0}}1=\frac{1}{4^z}\sum_{\substack{m=(4\lambda+1)(h+1)\\ \lambda,h\geq 0}}\left(\lambda+\frac{1}{4}\right)^{-z}(h+1)^{-z}=$$

$$=\frac{1}{4^z}\sum_{\lambda\geq 0}\left(\lambda+\frac{1}{4}\right)^{-z}\sum_{h\geq 0}(h+1)^{-z}=\frac{1}{4^z}\zeta\left(z,\frac{1}{4}\right)\zeta(z).$$

The computation for $\tau(m)$ is similar and yields

$$\sum_{m\geq 1}\frac{\tau(m)}{m^z}=\frac{1}{4^z}\zeta\left(z,\frac{3}{4}\right)\zeta(z). \quad □$$

THEOREM 2.12. *For all $m>0$ and $n\to\infty$ we have*

$$\sum_{m\geq 1}\left[(\sigma(m)-\tau(m))\operatorname{erfc}\left(\frac{m}{2\sqrt{n}}\right)\right]=\frac{1}{2}\sqrt{n\pi}-\frac{1}{4}+O(n^{-m}). \tag{2.20}$$

PROOF. By inversion of the Mellin transform of $\operatorname{erfc}(x)$ we get

$$\operatorname{erfc}(x)=\frac{1}{2\pi i}\frac{1}{\sqrt{\pi}}\int_{c-i\infty}^{c+i\infty}\frac{1}{z}\Gamma\left(\frac{z+1}{2}\right)x^{-z}\,dz, \qquad x>0,\quad c>0, \tag{2.21}$$

which can be found in [4, p. 325]. Thus (2.20) equals

$$\sum_{m\geq 1}(\sigma(m)-\tau(m))\frac{1}{2\pi i}\frac{1}{\sqrt{\pi}}\int_{c-i\infty}^{c+i\infty}\frac{1}{z}\Gamma\left(\frac{z+1}{2}\right)\left(\frac{m}{2\sqrt{n}}\right)^{-z}dz.$$

The sum may be placed inside the integral, since convergence is absolutely well behaved (cf. [8, p. 133]) and thus the last expression equals

$$\frac{1}{2\pi i}\frac{1}{\sqrt{\pi}}\int_{c-i\infty}^{c+i\infty}\frac{1}{z}\left(\frac{\sqrt{n}}{2}\right)^{z}\zeta(z)\left[\zeta\left(z,\frac{1}{4}\right)-\zeta\left(z,\frac{3}{4}\right)\right]\Gamma\left(\frac{z+1}{2}\right)dz. \tag{2.22}$$

By a well-known method it can be shown that the line of integration can be shifted to the left as far as we please if we only take the residues into account, yielding an error of $O(n^{-m})$ for all $m>0$. It is important to notice that, by cancellation of the poles of Hurwitz' zeta functions, $\zeta\left(z,\frac{1}{4}\right)-\zeta\left(z,\frac{3}{4}\right)$ is an entire function! At $z=1$ there is a simple pole with residue

$$\frac{\sqrt{n\pi}}{2}. \tag{2.23}$$

At $z=0$ there is a simple pole with residue

$$-\frac{1}{4}. \tag{2.24}$$

At $z=-(2k+1)$, $k\in\mathbf{N}_0$, the residues are zero, because

$$\begin{aligned}&\zeta\left(-(2k+1),\frac{1}{4}\right)-\zeta\left(-(2k+1),\frac{3}{4}\right)=\\ &=\frac{1}{2(k+1)}\left[-B_{2(k+2)}\left(\frac{1}{4}\right)+B_{2(k+2)}\left(1-\frac{1}{4}\right)\right]=0\end{aligned} \tag{2.25}$$

where $B_m(x)$ is the m-th Bernoulli polynomial (cf. [3, p. 49]). Summing up we get the desired result. □

THEOREM 2.13. *For $m>0$ and $n\to\infty$,*

$$\sum_{m\geq 1}(\sigma(m)-\tau(m))me^{-m^2/n}=\frac{n\pi}{8}+O(n^{-m}). \tag{2.26}$$

PROOF. Using the well-known formula

$$e^{-x}=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Gamma(z)x^{-z}\,dz, \qquad x>0,\quad c>0 \tag{2.27}$$

a similar method as in the proof of Theorem 2.12 (cf. e.g. [2]) yields the result. □

We want to summarize the results of this section in the following theorem.

THEOREM 2.14. *The expected maximal deviation of a path of length n, n even, is given by*

$$D_n=\sqrt{\frac{n\pi}{2}}-\frac{1}{2}+O(n^{-1/2+\varepsilon}) \quad \textit{for all} \quad \varepsilon>0 \quad \textit{and} \quad n\to\infty.$$

PROOF. The application of Theorems 2.12 and 2.13 to (2.17) yields the result. □

3. The expected maximal span

Let $\Psi_h(z)$ be the generating function whose n-th coefficient gives the number of paths of length n with maximal span $\leq h$. Such a path is shown in the following figure.

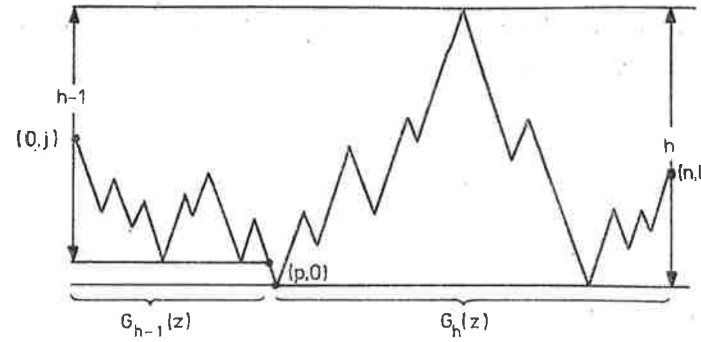

Fig. 2

To obtain an explicit expression for $\Psi_h(z)$ we proceed as follows: The minimal value of the ordinate of a path enumerated by $\Psi_h(z)$ is interpreted as the 0-level. Considering the point $P=(p, 0)$, where the minimal level is reached for the first time (in a-left-to-right-sense), the desired generating function can be obtained as the convolution of the two generating functions describing the left and the right part: For the right part we have to count the number of nonnegative paths with height $\leqq h$. For the left part we proceed from P to the left. If $p=0$, the contribution is 1; otherwise the first step leads to the point $(p-1, 1)$. For the number of the remaining paths we have to count the nonnegative paths of height $\leqq h-1$.

Using the generating function $G_h(z)$ of nonnegative paths with height $\leqq h$, defined in [6], we have

THEOREM 3.1.

$$\Psi_h(z) = (1+zG_{h-1}(z))\,G_h(z). \quad \square \tag{3.1}$$

The generating functions $G_h(z)$ are given by

$$G_h(z) = \frac{(1+v^2)(1-v^{h+1})}{(1-v)(1+v^{h+2})} \tag{3.2}$$

where the substitution $z=v/(1+v^2)$ was used. We would like to emphasize that (3.1) holds also for $h=0$, since $G_{-1}(z)=0$. Substituting (3.2) in (3.1) we get

$$\Psi_h(z) = \frac{(1+v^2)(1-v^{h+1})(1-v^{h+2})}{(1-v)^2(1+v^{h+1})(1+v^{h+2})}, \qquad h \geqq 0. \tag{3.3}$$

Let $d_{n,h}$ denote the number of paths of length n and maximal span $>h$ and $s_n = \sum_{h=0}^{n} d_{n,h}$.

THEOREM 3.2. *s_n is the coefficient of v^n in*

$$2\frac{(1+v^2)^n(1+v)v}{(1-v)^2}. \tag{3.4}$$

PROOF. The generating function $\chi_h(z)$ of all paths with maximal span $>h$ is the difference between the generating function of all paths $\left(\text{i.e. } \frac{1}{1-2z} = \frac{1+v^2}{(1-v)^2}\right)$ and $\Psi_h(z)$. Hence

$$\chi_h(z) = 2\frac{1+v^2}{(1-v)^2}\frac{(1+v)v^{h+1}}{(1+v^{h+1})(1+v^{h+2})}. \tag{3.5}$$

By Cauchy's integral formula we obtain

$$s_n = \sum_{h=0}^{n} d_{n,h} = \sum_{h=0}^{n} \frac{1}{2\pi i}\int^{(0_+)} \frac{dz}{z^{n+1}}\chi_h(z) =$$

$$= \frac{1}{2\pi i}\int^{(0_+)} \frac{dv}{v^{n+1}} 2\frac{(1+v)^2(1+v^2)^n}{(1-v)^2}\sum_{h=0}^{n}\left(\frac{1}{1+v^{h+2}} - \frac{1}{1+v^{h+1}}\right) =$$

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

which immediately leads to the desired result. $\square$

THEOREM 3.3.

$$s_{2n} = 2\left[(4n+1)\binom{2n-1}{n} - 2^{2n-1}\right],$$

$$s_{2n+1} = 2\left[4(2n+1)\binom{2n-1}{n} - 2^{2n}\right].$$

PROOF. This can be obtained by (3.4) using the following identities (cf. [11, p. 34]):

$$\sum_{k=0}^{n}\binom{2n}{n-k} = 2^{2n-1} + \binom{2n-1}{n}, \qquad \sum_{k=0}^{n}\binom{2n+1}{n-k} = 2^{2n},$$

$$\sum_{k=0}^{n} k\binom{2n}{n-k} = n\binom{2n-1}{n}, \qquad \sum_{k=0}^{n} k\binom{2n+1}{n-k} = (2n+1)\binom{2n-1}{n} - 2^{2n-1}. \quad \square$$

THEOREM 3.4. *The expected maximal span of a path of length n is given by*

$$\sqrt{\frac{8n}{\pi}} - 1 + O(n^{-1/2}), \qquad n \to \infty.$$

PROOF. Since the expected maximal span is $2^{-n}s_n$, we obtain the result by Theorem 3.3 and Stirling's formula. $\square$

4. The expected onesided height

Let $\Psi_{h,k;l}(z)$ be the generating function whose n-th coefficient corresponds to all paths with $-k \leqq a_i \leqq h$ for $0 \leqq i \leqq n$, $a_0=0$ and $a_n=l$; $\Psi_h(z)$ corresponds to all paths with $a_i \leqq h$ and $a_n=0$.

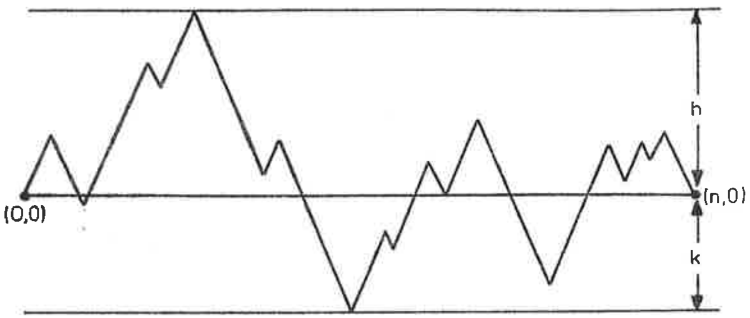

Fig. 3

THEOREM 4.1. *With* $z=v/(1+v^2)$,

$$\Psi_h(z) = \frac{1+v^2}{1-v^2}(1-v^{2h+2}). \tag{4.1}$$

PROOF. By a similar argument as in the proof of Theorem 2.1, we find that

$$\Psi_{h,k;0}(z) = \frac{a_{h-1}(z)a_{k-1}(z)}{a_{h+k}(z)} \tag{4.2}$$

with the determinants $a_i(z)$ of Theorem 2.1. Hence

$$\Psi_{h,k;0}(z) = \frac{1+v^2}{1-v^2}[1-v^{2h+2}]\frac{1-v^{2k+2}}{1-v^{2h+2k+4}}. \tag{4.3}$$

Since $\Psi_h(z)=\lim_{k\to\infty}\Psi_{h,k;0}(z)$ we have the result for small values of v (and thus small values of z) and by continuation for all values of z in the circle of convergence of $\Psi_h(z)$. □

Let s_n denote the sum of heights of all paths of length $2n$.

THEOREM 4.2.

$$s_n = 2^{2n-1}-\binom{2n-1}{n}. \tag{4.4}$$

PROOF. Let $D(z)=(1-4z^2)^{-1/2}=\sum_{n\geqq 0}\binom{2n}{n}z^{2n}$ be the generating function of all paths of length $2n$. Then, as usual, s_n is the coefficient of z^{2n} in

$$\sum_{h\geqq 0}(D(z)-\Psi_h(z)) =$$

$$= \sum_{h\geqq 0}\left(\frac{1+v^2}{1-v^2}-\frac{1+v^2}{1-v^2}(1-v^{2h+2})\right) = \frac{1+v^2}{1-v^2}\sum_{h\geqq 0}v^{2h+2} = v^2\frac{1+v^2}{(1-v^2)^2}. \tag{4.5}$$

$$s_n = \frac{1}{2\pi i}\int^{(0+)}\frac{dz}{z^{2n+1}}v^2\frac{1+v^2}{(1-v^2)^2} = \frac{1}{2\pi i}\int^{(0+)}\frac{dv}{v^{2n+1}}\frac{v^2(1+v^2)^{2n}}{1-v^2}.$$

Hence s_n is the coefficient of u^n in

$$\frac{u(1+u)^{2n}}{1-u},$$

which is

$$\sum_{\lambda\geqq 1}\binom{2n}{n-\lambda}. \quad □ \tag{4.6}$$

COROLLARY 4.3. *The expected onesided height of a path of length 2n is given by*

$$\frac{1}{2}\sqrt{\pi n}-\frac{1}{2}+O(n^{-1/2}) \quad \textit{for} \quad n\to\infty. \tag{4.7}$$

PROOF. Apply Stirling's approximation formula to $\binom{2n}{n}^{-1}s_n$. □

5. Concluding remarks

We would like to mention that Theorem 2.14 also holds true for n odd. This could be shown by additionally considering the case n odd in the derivation of Theorem 2.14. But it suffices for our purposes to observe that the expected maximal deviation is a strictly increasing function of path length. Now — since the difference $D_{2n+2}-D_{2n}$ is of order $O(n^{-1/2})$ — the validity of Theorem 2.14 for all n immediately follows.

Distinguishing the three components of the approximation error dealt with in the proof of Theorem 2.9, a good balance of their order was achieved by taking the correction terms into account in Theorem 2.9 (ii) and (iii). By that means we were able to lower the order of the error term in Theorem 2.14 from $O(n^\varepsilon)$ to $O(n^{-1/2+\varepsilon})$ and thereby the absolute term could be preserved. In principle better approximations could be obtained by use of Euler's summation formula. However, it can be seen from the following table that the asymptotic formulae derived in Section 2 show an accuracy meeting most practical requirements even for small n.

Furthermore, a point of some methodical interest should be emphasized: Though the three problems treated in Sections 2 to 4 seem to be very akin, in the study of their asymptotic behaviour different methods had to be used.

In Section 3 and 4 the exact formulae were reduced to a relatively simple form by the use of some combinatorial identities. To derive the corresponding asymptotic formulae essentially Stirling's formula had to be applied.

n \ D_n	exactly (Theorem 2.8)	asymptotically (Theorem 2.14)
10	3.53	3.46
20	5.15	5.11
30	6.40	6.36
40	7.46	7.43
50	8.39	8.36
60	9.24	9.21
70	10.01	9.99
80	10.73	10.71
90	11.41	11.39
100	12.05	12.03

For the study of the asymptotic behaviour of the expected maximal deviation treated in Section 2, the so-called *Γ-function method,* used e.g. in [2], had to be modified: instead of the exponential function the complement of the error function was used. In addition to Riemann's zeta-function also Hurwitz' zeta-function came into play.

REFERENCES

[1] Apostol, T. M., *Introduction to Analytic Number Theory*, Springer, New York—Heidelberg—Berlin, 1976. *MR* **55** # 7892.

[2] de Bruijn, N. G., Knuth, D. E. and Rice, S. O., The average height of planted plane trees, in: *Graph theory and computing*, R. C. Read, Ed., 15—22. Academic Press, New York—London, 1972. *MR* **48** # 8280.

[3] Comtet, L., *Advanced Combinatorics*, Cloth Edition, D. Reidel, Dordrecht, 1974. *MR* **57** # 124.

[4] Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F. G., *Tables of Integral Transform, Vol.* 1, McGraw-Hill, New York—Toronto—London, 1954. *MR* **15**—868.

[5] Feller, W., *An introduction to probability theory and its applications*, Vol. 1, 3rd ed., J. Wiley, New York, 1968. *MR* **37** # 3604.

[6] Kemp, R., On the average depth of a prefix of the Dycklanguage D_1, *Discrete Math.* **36** (1981), 155—170.

[7] Kirschenhofer, P. and Prodinger, H., On the average hyperoscillations of planted plane trees, *Combinatorica* **2** (1982), 177—186.

[8] Knuth, D. E., *The Art of Computer Programming*, Vol. 3, Addison-Wesley, Reading, Massachusetts, 1973. *MR* **56** # 4281.

[9] Panny, W., The expected depth of a pushdown store with three operations and symmetrical probabilities, *J. Comb. Inf. Syst. Sci.* **7** (1982), 38—47.

[10] Prodinger, H., The average height of a stack where three operations are allowed and some related problems, *J. Combin. Inform. System Sci.* **5** (1980), 287—304. *MR* **82 m**: 05013.

[11] Riordan, J., *Combinatorial Identities*, J. Wiley, New York, 1968. *MR* **38** # 53.

[12] Whittaker, E. T. and Watson, G. N., *A course of modern analysis*, 4th ed., Cambridge, University Press, 1927.

(Received January 28, 1982)

INSTITUT FÜR STATISTIK
WIRTSCHAFTSUNIVERSITÄT WIEN
AUGASSE 2—6
A—1090 WIEN

INSTITUT FÜR ALGEBRA UND
DISKRETE MATHEMATIK
TECHNISCHE UNIVERSITÄT
WIEDNER HAUPTSTRASSE 8—10
A—1040 WIEN
AUSTRIA

86-1723 — Szegedi Nyomda — Felelős vezető: Surányi Tibor igazgató