Asymptotic properties of some underdiagonal walks generation algorithms

G. Louchard *

August 14, 2000

Abstract

Using classical properties of Random walks and Brownian motion, we derive the asymptotic distribution of some characteristics of an algorithm proposed by Barcucci et al. [2] to generate underdiagonal walks of size n. The main parameters of interest are the cost of the algorithm (in terms of calls to a random generator), the height of the walk at its last step and the maximum of the walk. We obtain various forms of probability generating functions, Laplace transforms of densities, continuous distribution functions, asymptotic densities and discrete stationary distributions.

1 Introduction

1.1 Underdiagonal walks and their generation

In the litterature, several papers have been devoted to the analyzis of one-dimensional walks made of three kinds of unitary steps: right (r), left (l) and in-place (s). Particular attention has been given to walks that begin at the origin 0 and never go to its left. This means that in each subwalk which begings at the origin 0, the number of right-hand steps must be greater than, or equal to, the number of left-hand steps. This kind of walk has been very thoroughly studied because these walks correspond biunivocally to single-rooted directed animals.

The walks on the line can also be represented two dimensionally by recording the time on the abscissa and the moves on the ordinates. As a result, the elementary steps corresponding to the right, left and in-place steps are north-east, south-east and east, respectively. In this case, the condition that a one-dimensional walk must not go to the left of the origin means that it must not go under the x axis.

Another way of representing this kind of walk on a plane is to consider the east, north, and north-east steps as elementary steps corresponding to r, l, s. In this case, the condition that

*Université Libre de Bruxelles, Département d'Informatique - C.P. 212, Campus de la Plaine, Boulevard du Triomphe, B-1050 Bruxelles (Belgium); email: louchard@ulb.ac.be

the one-dimensional walk must not go to the left of 0 means that the walk must not go over the $y = x$ diagonal. This is why these walks are also called *underdiagonal* walks.

In [3], Barcucci et al. introduced a linear algorithm for the random generation of single-rooted directed animals, whose first step is to generate an underdiagonal walk. In [2] they examine a more general class of underdiagonal walks made up of several kinds of east, north and north-east steps and they introduce an algorithm that generates them randomly. This algorithm is shown to be linear when the number of east steps is greater than, or equal to, the number of north steps.

The authors examine underdiagonal walks in which a kinds of east steps, b kinds of north-steps and c kinds of north steps can be used. This type of walk is usually called *coloured walk*. Walks made up of n steps (n-long walks or n-walks) can be codified by n-long words (n-words) on the alphabet $\mathcal{A} = \{x_1, x_2, \cdots, x_a, z_1, \cdots, z_b, y_1, \cdots, y_c\}$, in which the x_i, z_i and y_i represent the east, north-east and north steps.

For a walk to be underdiagonal, the number of north steps in each subwalk beginning at the origin must be less than, or equal to, the number of east steps. Consequently, if $W \in \mathcal{A}^*$ is a word that codifies an underdiagonal walk, we obtain

$$\sum_{k=1}^{c} |w'|_{y_k} \leq \sum_{k=1}^{a} |w'|_{x_k}$$

for each w' prefix of w.

In order to generate a random underdiagonal n-walk, one has to generate a random n-word. The authors propose the following algorithm in [2].

(1) The letter of the word are generated one after another by taking them at random out of $\{x_1, x_2, \cdots, x_a, z_1, \cdots, z_b, y_1, \cdots, y_c\}$.

(2) The difference $\sum_{i=1}^{a} |w|_{x_i} - \sum_{j=1}^{c} |w|_{y_j}$ is taken into account.

(3) If this difference becomes negative before generating n letters, the prefix generated up to then is discarded and one starts from the beginning again.

(4) If n letters are generated before the difference is less than zero, they constitute the word desired.

1.2 Probabilistic interpretation

The walks defined in Sec. 1.1 can be seen as random walks (RW): let X_i be individual steps (identically independent distributed random variables) and let $Y(n) := \sum_{i=1}^{n} X_i$ be the associated RW. In our case, we have $p = \Pr(X_i = 1) = a/d, r = \Pr(X_i = 0) = b/d, q = \Pr(X_i = -1) = c/d (d := a + b + c)$ and we consider a reflected RW (to take action (3) of the algorithm into account) starting from 1. Test (4) of the algorithm amounts to wait for the first Meander of duration $\geq n$.

A Meander $V(\cdot)$ of duration m is defined as a RW starting from 0 and conditioned to stay strictly positive for $i \leq m$. Similarly, a RW starting from 0, returning to 0 at time m, and staying strictly positive for $0 < i < m$ is called an Excursion $Z(\cdot)$ of duration m. Note that the RW is transient if $p > q$, null recurrent if $p = q$, positive recurrent if $p < q$ (see Spitzer [12]).

To summarize, the input to the algorithm is (a, b, c) and the output is a walk $V(\cdot)$ of length n, such that $V(i) > 0, i = 1 \cdot \cdot n$.

The cost c_n of the algorithm is defined as the number of steps (excluding the steps from 0 to 1 or 0 and the last n steps), necessary to obtain a Meander of duration n. c_n is actually the number of generated letters necessary for obtaining a n-word, i.e. the number of calls to a random number generator (excluding the last n steps).

The deviation δ_n is defined by the height of the Meander at its last step. c_n is related to the number of Excursions before the first useful Meander, i.e. to the number R of returns to the origin before a Meander of duration n.

The results of [2] can be summarized in the following Table 1 where M := mean, VAR := variance, and we keep only the dominant term of each expression ($n \to \infty$).

	c_n		δ_n
	M	VAR	M
$p > q$	$\frac{q}{(p-q)^2}$	$\frac{q}{(p-q)^4}[2p - p^2 + 2pq + q - q^2]$	$(p-q)n$
$p = q$	n	$\frac{4n^2}{3}$	$\sqrt{\frac{\pi n}{2}}\sigma$
$p < q$	$\frac{2p\sqrt{\pi}(\sqrt{q}-\sqrt{p})^2}{(pq)^{1/4}(r+2\sqrt{pq})^{3/2}(q-p)} \cdot$ $\cdot z^{*n} n^{3/2}$		$\frac{\sqrt{q}+\sqrt{p}}{\sqrt{q}-\sqrt{p}}$ (+)

Table 1 : dominant term of some characteristics of c_n and δ_n.

Here, $z^* = \frac{1}{r+2\sqrt{pq}} > 1$. (+): The expression $\frac{2\sqrt{p}}{\sqrt{q}-\sqrt{p}}$ given in [2] yields actually $M - 1$, since our RW is one unit higher than the walk generated in [2].

1.3 Asymptotic properties

In the field of analysis of algorithms, the moments of a cost distribution are usually the first steps in the complete study: asymptotic distributions of costs and related random variables (RV) are more informative and shed more light on their stochastic behaviour.

The purpose of the present paper is to analyze the asymptotic distributions of c_n and δ_n, $n \to \infty$.

Apart from rederiving easily the main terms as given in Table 1, we shall obtain various forms of asymptotic characteristics, such as probability generating functions (PGF), Laplace transforms of densities ($\mathcal{L}_\alpha$), continuous distribution functions (DF), asymptotic densities (Gaussian, exponential, extreme value, Jacobi or Rayleigh), discrete stationary distribution.

So we have a complete asymptotic probabilistic analysis of the two essential parameters, c_n and δ_n of the generation algorithm proposed in [2].

Of course all asymptotic moments (ordinary and central) can be immediately deduced from our expressions, which can be summarized in Table 2 and 3.

	c_n
$p>q$	PGF $(c_n) \sim$ $\frac{1-q/p}{1-\lambda_2(z)}$ (T1)
$p=q$	$E[\exp(-\alpha c_n/n)] \sim$ $\frac{1}{e^{-\alpha}+\sqrt{\pi\alpha}\ \mathrm{erf}(\sqrt{\alpha})}$ (T2)
$p<q$	DF $(x=\frac{c_n}{M}) \sim$ (exponential RV) $1-e^{-x}$

Table 2: asymptotic characteristics of c_n .

	δ_n	
	Asymptotic characteristics	VAR
$p>q$	$\frac{\delta_n-(p-q)n}{\sqrt{n}\sigma} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1)$ (Gaussian RV)	$n\sigma^2$
$p=q$	$v=\delta_n/(\sqrt{n}\sigma) \overset{\mathcal{D}}{\sim}$ Brownian Meander, with density $ve^{-v^2/2}$ (T3) (Rayleigh density)	$[2-\frac{\pi}{2}]n\sigma^2$
$p<q$	Stationary distribution : $\pi(i)=\frac{P(1)q}{\sqrt{pq}}\left[\frac{i}{\mu_1^i}\right]$ (T4)	$\frac{2\sqrt{pq}}{p+q-2\sqrt{pq}}$

Table 3: asymptotic characteristics of δ_n.

with

$$
\begin{aligned}
\mathrm{VAR}(c_n) &\sim E^2(c_n)=M^2,\ \text{if } p<q \\
\sigma^2 &= p+q-(p-q)^2 \\
\lambda_2(z) &= \frac{(1-rz)-\sqrt{(1-rz)^2-4pqz^2}}{2pz} \qquad (1) \\
P(1) &= \frac{p+q-2\sqrt{pq}}{q} \\
\mu_1 &= \sqrt{\frac{q}{p}}>1 \\
\mathcal{L}_\alpha[\mathrm{T3}] &= \frac{1}{2}e^{\alpha^2/2}[2e^{-\alpha^2/2}-\sqrt{2\pi}\alpha\ \mathrm{erfc}(\alpha/\sqrt{2})] \qquad (2) \\
\mathrm{PGF}\ [\mathrm{T4}] &= qP(1)z/(pz^2+z(qP(1)-p-q)+q) \qquad (3)
\end{aligned}
$$

1.4 The Meander maximum

We have also analyzed the maximum of the Meander : $MM=\max V(i),\ i\in[0,n]$ and its position $I\in[0,n]$. Our results are summarized in Table 4.

$p > q$	$\frac{MM-(p-q)n}{\sqrt{n}\sigma} \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0,1)$
$p = q$	$DF(\mathcal{M}) \sim F_1(m^2/2)$ Probability density $(\mathcal{M},\mathcal{T}) \sim \frac{\pi^2}{m^2}\frac{f_1(\frac{\pi^2}{2m^2}t)}{2m} \cdot \frac{F_1[m^2/(2(1-t))]}{\sqrt{1-t}}$
$p < q$	$\Pr\{MM < [ln\ n + ln\ ln\ n - ln\ ln\ \mu_1 + \eta - ln\ C_6]/ln\ \mu_1\} \sim e^{-e^{-\eta}}$ (extreme value distribution) (T5)

Table 4: asymptotic distribution of MM and I

where

$$\begin{aligned} \mathcal{M} &= \frac{MM}{\sigma\sqrt{n}}, \text{with value m} \\ \mathcal{T} &= \frac{I}{n}, \text{with value t} \\ F_1(x) &= \sum_{i=-\infty}^{+\infty} (-1)^i e^{-i^2 x} \end{aligned}$$

with Jacobi density $f_1(x)$ such that $\mathcal{L}_\alpha(f_1) = \frac{\pi\sqrt{\alpha}}{sh(\pi\sqrt{\alpha})}$. f_1 is related to the third Jacobi θ function (see Louchard [9])

$$C_6 := \frac{\sqrt{pq}}{P(1)q(q-p)}.$$

1.5 Notations and tools

The next 3 sections cover the 3 cases : $p < q, p > q, p = q$. The last section is devoted to the Meander maximum analyzis.

The following notations will be used in the sequel :

- $\stackrel{\mathcal{D}}{\sim}$: convergence in distribution, for $n \to \infty$
- $\Longrightarrow_{m\to\infty}$: weak convergence of random functions in the space of all right continuous functions having left limits with values in R^2 and endowed with the Skorohod metric d_0 (see Billingsley [4], Ch. III.)
- $\mathcal{N}(M,V)$:= the Normal (or Gaussian) random variable with mean M and variance V
- Brownian Motion (BM) := Markovian Gaussian process, with mean $E[BM(t)] = 0$, variance VAR $[BM(t)] = t$ and covariance $E[BM(s)BM(t)] = s(s \leq t)$: see Ito and Mc Kean [7]
- Brownian meander: equivalent for the BM of the classical meander
- PGF := probability generating function: $\sum_0^\infty p_i z^i$ of the discrete probability distribution p_i
- $\mathcal{L}_\alpha$:= Laplace transforms of the density $f(x)$: $\int_0^\infty e^{-\alpha x} f(x)dx$

The computer algebra system MAPLE has been quite useful for detailed computations and simulations.

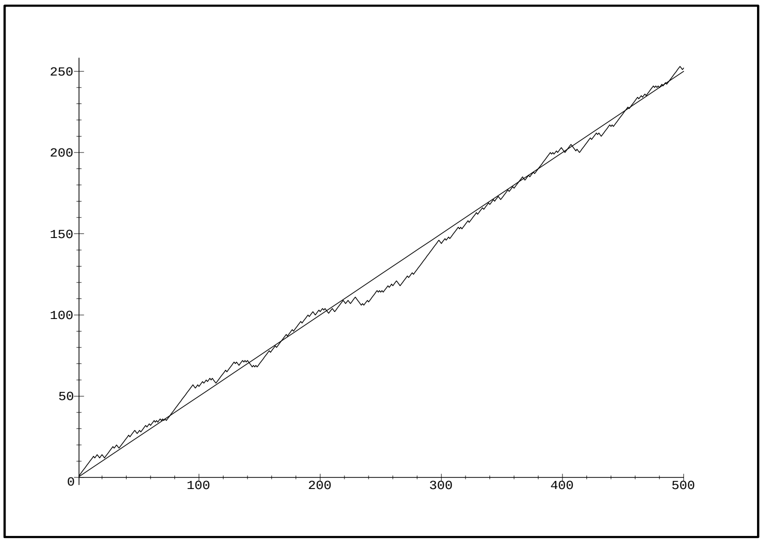

Figure 1: RW $Y(i), n = 500$

2 Case $p > q$

It is known (see for instance Louchard and Schott [8] Thm 2.3) that $\frac{Y([mt])-m(p-q)t}{\sqrt{m}\sigma} \Rightarrow BM(t)$, $m \to \infty$ where BM is the classical Brownian Motion. The asymptotic density of δ_n is now immediate as δ_n is asymptotically equivalent to the BM at $t = 1$. $c_n = \mathcal{O}(1)$ as we shall see and we must derive its asymptotic distribution. Notice that c_n is asymptotically equivalent to the last leaving time L from the origin for $Y(\cdot)$ (excluding the step from 0 to 1 or 0). The PGF of the hitting time to 0, T_0, starting from 1, is well known (see Feller [6] p. 496, ex. 12, where we now take $r > 0$ into account): this is given by $\lambda_2(z)$ (see (1)). Moreover the hitting probability is given by $\lambda_2(1) = \frac{q}{p}$. Hence the PGF for L is immediately given by PGF (L) $:= (1 - \frac{q}{p})/[1 - \frac{q}{p}[\lambda_2(z) \cdot \frac{p}{q}]]$, i.e., (T1).

All asymptotic central moments of c_n are immediately derived from (T1): just replace z by e^w and expand $e^{-wM}\cdot$ PGF (L) into w (this is trivial for MAPLE).

A typical RW for $n = 500$ and $p = \cdot 75$, $r = 0$, is given in Fig. 1, together with the shift $(p - q)i$. The corresponding normalized RW is given in Fig. 2

The behaviour of the RW during the first $\mathcal{O}(1)$ steps is illustrated by Fig. 3 (this is a typical zoom of the trajectory near the origin)

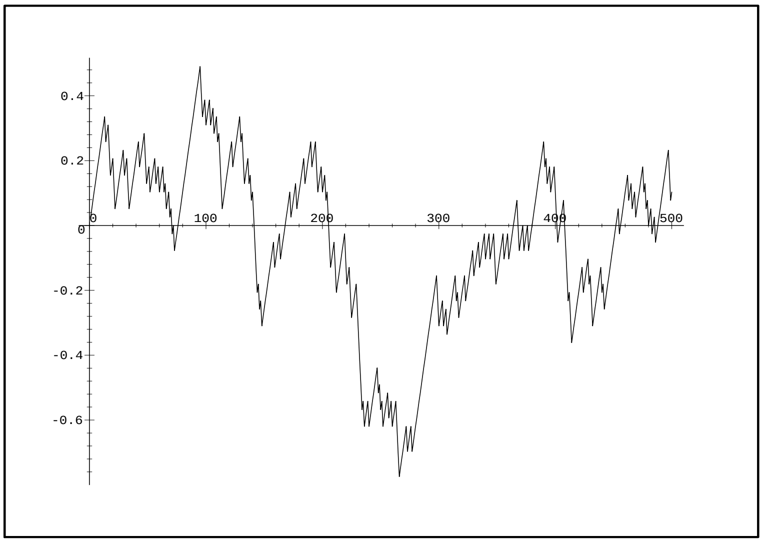

Figure 2: $\frac{Y(i)-(p-q)i}{\sqrt{n}\sigma}$

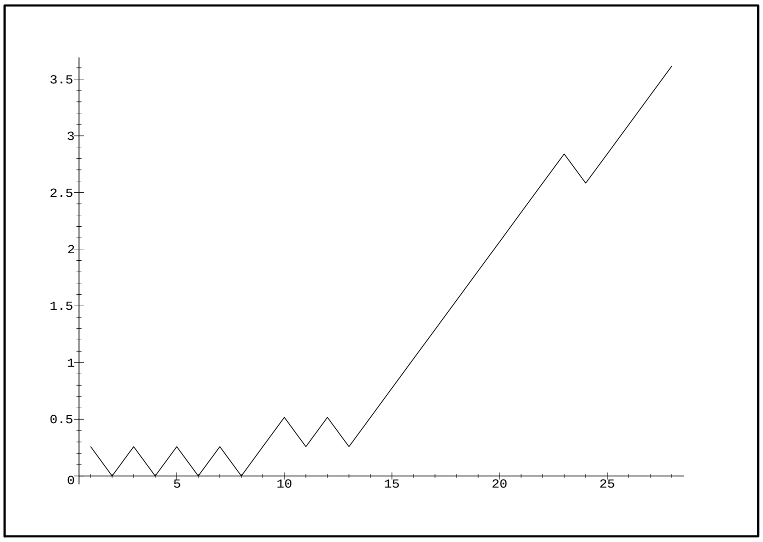

Figure 3: $Y(i)(i = 0 \cdot \cdot 28)$

3 Case $p < q$

3.1 Analysis of c_n

We will proceed as in Louchard and Schott [8]. Let us first obtain the probability

$$\varphi_n := P_1[T_0 \leq n].$$

where again T_0 is the hitting time from 1 to 0.

The GF of $\theta := 1 - \varphi_n$ is given by

$$\psi(z) := \frac{1 - \lambda_2(z)}{1 - z}.$$

Let $z - 1 = (q - p)\eta$. We check that $\lambda_2 \sim 1 + \eta + \mathcal{O}(\eta^2)$. 1 is not a singularity of $\psi(\cdot)$. So we look for the dominant algebraic singularity (of smallest module) of $\psi(\cdot)$. This is given by $z^* = \frac{1}{r+2\sqrt{pq}} > 1$. So we put $z = z^* - \epsilon$ and expand $\psi(\cdot)$ into ϵ. This gives a Puiseux series

$$\psi(z) \sim C_2 + C_3\sqrt{\epsilon} + \mathcal{O}(\epsilon) \sim C_2 + C_4\sqrt{1 - \frac{z}{z^*}} + \mathcal{O}(\epsilon),$$

with $C_3 := \frac{(1-p-q+2\sqrt{pq})^2 q^{1/4}}{p^{3/4}(2\sqrt{pq}-p-q)}$ and $C_4 := \sqrt{z^*}C_3$. Darboux's Thm. immediately leads to

$$\theta := 1 - \varphi_n = [z^n]\psi(z) \sim \frac{C_4}{\Gamma(-1/2)} \frac{1}{z^{*n} n^{3/2}}$$

The PGF for the last leaving time from 0 is now asymptotically given by

$$\frac{\theta}{1 - (1 - \theta)\bar{\lambda}_2(z)} \tag{4}$$

where $\bar{\lambda}_2(z)$ is the PGF of T_0, given that $T_0 < n$.

But, as in [8] B.2, it is easy to check that

- we can use λ_2 instead of $\bar{\lambda}_2$.
- λ_2 contributes to (4) only through its first approximation $\lambda_2 \sim 1 + \eta$.

Proceeding exactly as in [8] B.1, (4) readily becomes

$$\text{PGF (L)} \sim \frac{-\theta}{\eta - \theta} = \frac{-\epsilon_1}{z - z_1}, \qquad \text{when } \eta \to 0$$

with $\epsilon_1 := (q-p)\theta, z_1 := 1 + \epsilon_1$. This leads to the asymptotic exponential DF : $1 - e^{-c_n/M}$ with $M \sim 1/\epsilon_1$.

3.2 Analysis of δ_n

As $p < q$, it is clear that the Meander possess a stationary distribution. To derive this distribution, let us first write down the Meander probabilities $P^*_{m+1}(i)$ at time $m+1$, given that the (stationary) probability is given by $P(i)$ at time m. We have

$$\begin{aligned} P^*_{m+1}(i) &= P(i-1)p + P(i)r + P(i+1)q, \quad \text{for } i \geq 2 \\ P^*_{m+1}(1) &= P(1)r + P(2)q. \end{aligned}$$

We have lost a probability mass $P(1)q$. Normalizing P^*_{m+1} by $1 - P(1)q$ leads to the difference equation:

$$\begin{aligned} (1 - P(1)q)P(i) &= P(i-1)p + P(i)r + P(i+1)q, \qquad \text{for } i \geq 2 \\ (1 - P(1)q)P(1) &= P(1)r + P(2)q \\ P(0) &= 0 \end{aligned} \tag{5}$$

It may be checked that (T4) is the solution of this equation.

To fix $P(1)$, note first that the hitting probability $P_1[T_0 \sim m+1]$ is asymptotically given by

$$\frac{C_4}{\Gamma(-1/2)m^{3/2}}\left[\frac{1}{z^{*m}} - \frac{1}{z^{*(m+1)}}\right] = \frac{z^*-1}{z^*}\,\frac{C_4}{\Gamma(-1/2)}\,\frac{1}{m^{3/2}z^{*m}}$$

(This can also be derived from a λ_2 expansion). On the other way, this is also given by

$$\sim \frac{C_4}{\Gamma(-1/2)m^{3/2}z^{*m}} \cdot P(1)q$$

hence

$$P(1) = \frac{p+q-2\sqrt{pq}}{q}.$$

To analyze the moment of δ_n, it is easier to start from the PGF related to (5). This readily leads to (3). Again replace z by e^w and expand $e^{-wM}\cdot$ PGF (δ_n) into w. All central moments are immediately available.

A typical RW for $n = 40$ and $p = \cdot 25$, $r = 0$, is given in Fig. 4. We see that more than 12.000 steps are necessary to reach the meander. The behaviour of the RW during the last 40 steps is illustrated in Fig. 5.

4 Case $p = q$

4.1 Analysis of c_n

Proceeding as in case $p < q$, we readily obtain (the dominant singularity is now $z^* = 1$)

$$\theta := 1 - \varphi_n := P_1[T_0 > n] \sim \frac{\sqrt{2}}{\sqrt{\pi}n^{1/2}\sigma}. \tag{6}$$

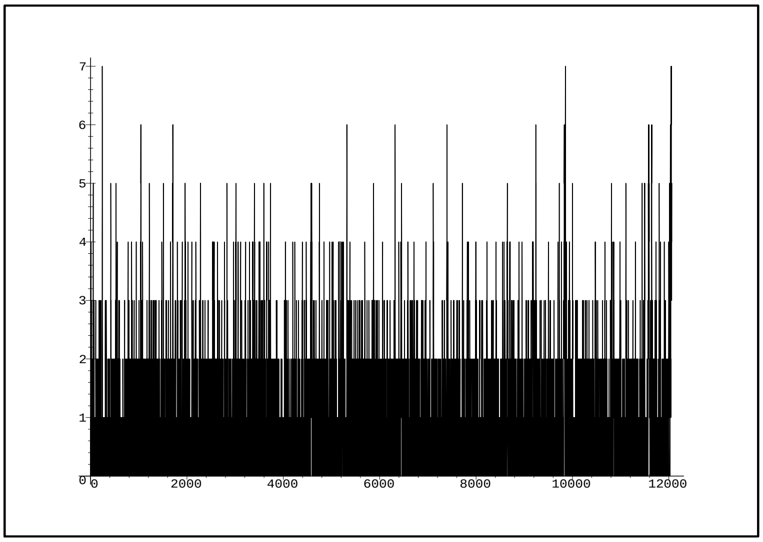

Figure 4: $Y(i), n = 40$

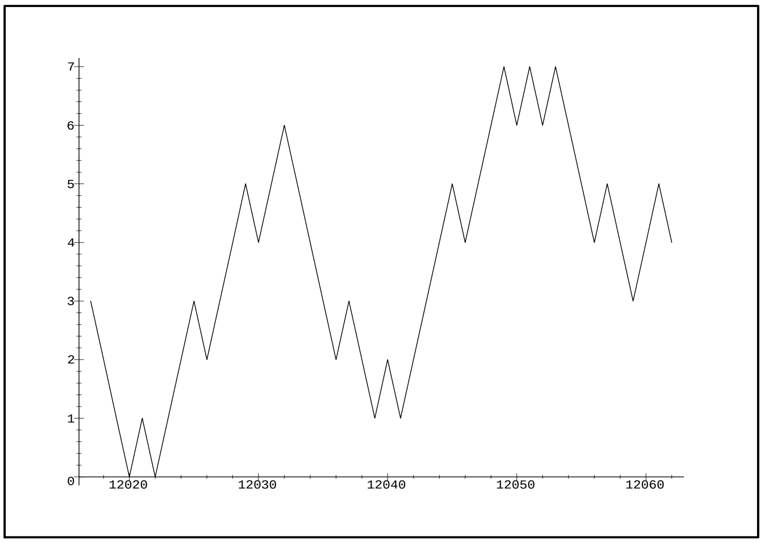

Figure 5: $Y(i), n = 40$, last 45 steps

Note that by Spitzer [12] p. 378, we know that

$$P_0[T_0 > n] \sim \frac{\sqrt{2}\sigma}{\sqrt{\pi}n^{1/2}}.$$

But $P_0[T_0 > n] = 2pP_1[T_0 > n] = \sigma^2 P_1[T_0 > n]$. The case $p = q$ is of course related to Motzkin paths (see [2]). We have

$$P_1[T_0 = n] \sim \frac{1}{\sqrt{2\pi}\sigma n^{3/2}}. \tag{7}$$

(From (6) or from a λ_2 expansion). Of course $E_1[T_0] = \infty$ (the RW is null recurrent) but here, the Excursions before the first useful Meander are of limited duration $< n$.

Let us first heuristically derive (T2) as follows. The number R of returns to 0 before a length n meander is geometrically distributed with parameter φ_n. Moments of T_0 are asymptotically given from (7) as

$$E_1[T_0^k] \sim \int_0^n \frac{1}{\sigma\sqrt{2\pi}u^{3/2}} u^k du, k \geq 1$$

So

$$E[e^{-\alpha c_n/n}] \sim \frac{\theta}{1 - (1-\theta)\left[1 + \int_0^n (e^{-\alpha u/n} - 1)\frac{du}{\sigma\sqrt{2\pi u^3}}\right]}.$$

Set $u = n\ell$, this leads to

$$\sim \frac{1}{1 - \frac{\sqrt{\pi}}{\sqrt{2}}\int_0^1 (e^{-\alpha\ell} - 1)\frac{d\ell}{\sqrt{2\pi\ell^3}}}. \tag{8}$$

To compute the integral, note that

$$-\partial_\alpha \int_0^1 (e^{-\alpha\ell} - 1)\frac{d\ell}{2\sqrt{l^3}} = \int_0^1 e^{-\alpha\ell}\frac{d\ell}{2\sqrt{\ell}} = \frac{\sqrt{\pi}}{2\sqrt{\alpha}}\operatorname{erf}(\sqrt{\alpha}).$$

But

$$\partial_\alpha\{\sqrt{\pi}(\sqrt{\alpha}\operatorname{erf}(\sqrt{\alpha}) + e^{-\alpha}/\sqrt{\pi})\} = \frac{\sqrt{\pi}}{2\sqrt{\alpha}}\operatorname{erf}(\sqrt{\alpha}).$$

So,

$$-\int_0^1 (e^{-\alpha\ell} - 1)\frac{d\ell}{2\sqrt{\ell^3}} = \sqrt{\alpha\pi}\operatorname{erf}(\sqrt{\alpha}) + e^{-\alpha} + C_5$$

and C_5 is easily checked to be given by - 1, hence (T2). All asymptotic moments of c_n/n are immediate from (T2).

Let us now derive (T2) rigourously. We use the same techniques as in Louchard [9] Sec.4. First, from (6) the number of returns to the origin, R, is such that $\frac{\sqrt{2}}{\sigma\sqrt{\pi n}} \cdot R$ is asymptotically distributed as an exponential RV. Now,

$$\frac{Y([nt])}{\sigma\sqrt{n}} \Rightarrow \mathrm{RBM}(t)$$

where RBM (t) is a reflected Brownian Motion $x^+(\cdot)$. Excursions and Meandering processes can be rigourously defined for a BM: see Chung [5]. The local time τ^+ at the origin, up to R returns to 0 is given by

$$\tau^+ \sim \frac{(1+\frac{r}{1-r})\cdot\frac{1}{n}\cdot R}{1/(\sqrt{n}\sigma)} = \frac{R}{\sigma\sqrt{n}}$$

So $\tau^+\sqrt{\frac{2}{\pi}}$ is asymptotically distributed as an exponential RV.

Let $\tau^+(s)$ be the local time τ^+ up to time s and

$$\tau^{-1}(b) := \inf(s : \tau^+(s) = b)$$

then

$$\tau^{-1}(b) = \int_0^\infty \ell p([0,b] \times d\ell)$$

where $p(db \times d\ell)$ is the Poisson measure with mean $db \times (d\ell/\sqrt{2\pi\ell^3})$ (see Ito and McKean [7], §.1.7). This decomposition ties the flat stretches of τ^+ with the open intervals $z_n (n \geq 1)$ of the complement of the set $z^+ \equiv (t : x^+(t) = 0)$. The Excursions $Z_n(u) \equiv [x^+(u) : u \in z_n]$ are independent and depend only on each length $|z_n|$. This leads to

$$E[\exp(-\alpha\tau^{-1}(b))] = e^{-\sqrt{2\alpha}b}.$$

But, here, we must limit ourselves to Brownian Excursions of maximum duration 1 (otherwize, we get a meander of length $\geq n$). So,

$$E[e^{-\alpha\tau^{-1}(b)}] = \exp\left[b\int_0^1 (e^{-\alpha\ell}-1)\frac{d\ell}{\sqrt{2\pi\ell^3}}\right].$$

And, finally,

$$E[e^{-\alpha c_n/n}] \sim \int_0^\infty e^{-b\sqrt{\frac{2}{\pi}}} \exp\left[b\int_0^1 (e^{-\alpha\ell}-1)\frac{d\ell}{\sqrt{2\pi\ell^3}}\right] db\sqrt{\frac{2}{\pi}}$$

which is equivalent to (8). Note that $\tau^{-1}(b)$ takes steps from 0 to 1 into account, but, this is an $O(\sqrt{n})/n$ contribution, negligeable w.r.t. the RV characterized by (T2).

We have simulated a sample of $N = 2000$ RW, with $n = 2000$, $p = 0.5$. The observed distributions of R (returns to 0) and $e^{-\alpha c_n/n} (\alpha = 0 \cdot\cdot 5)$ fit remarkably with the limiting distributions.

4.2 Analysis of δ_n

Let us now turn to the normalized Meander $\delta_n/(\sqrt{n}\sigma)$. The density of the classical BM Meander is well known: this is given by the Rayleigh expression

$$\frac{v}{t}e^{-v^2/(2t)}$$

and here $t = 1$.

$\mathcal{L}_\alpha(ve^{-v^2/2})$ is given by (2), from which all moments are easy to compute.

A typical normalized RW for $n = 500$, $p = 0.5$, is given in Fig. 6

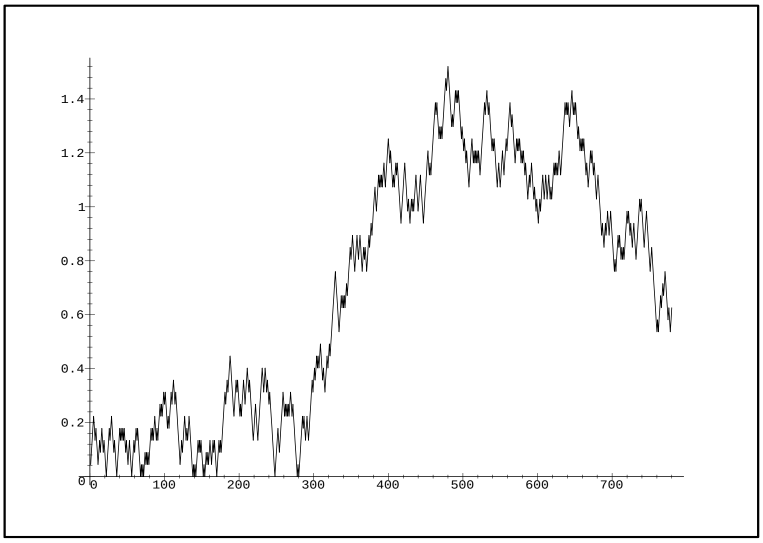

Figure 6: $\frac{Y(i)}{\sqrt{n}\sigma}$

5 The Meander Maximum

5.1 Case $p > q$

As the drift $(p-q)$ is strictly positive, it is clear that the maximum is asymptotically obtained for $I = n$ and its value is asymptotically given by a Gaussian.

5.2 Case $p = q$

The joint probability of the maximum and its position for a BM Meander on $[0,1]$ is given in Louchard [9], (34). The maximum position is almost surely unique.

5.3 Case $p < q$

We shall proceed as in Louchard [11], i.e. with the help of the Poisson Clumping Heuristic (see Aldous [1] B16). The RW, conditioned on no-absorption, is irreducible, positive and recurrent. Then, for $K >> 1$, the hitting time T_K is such that

$$\Pr[T_K \geq n] \sim e^{-n/E[T_K]} \tag{9}$$

and $E[T_K] \sim E(C)/\pi(K)$ where $\pi(\cdot)$ is the stationary distribution given by (T4) and C is the sojourn time in K during a clump of nearby visits to K. We have $\pi(K) \sim K\ C_5/\mu_1^K$, with $C_5 := \frac{P_1 q}{\sqrt{pq}}$ and $E(C) = 1 + pE(C) + r\ E(C) + q \cdot \frac{p}{q} E(C)$, i.e. $E(C) = \frac{1}{q-p}$. So $E[T_K] \sim \frac{C_6 \mu_1^K}{K}$, with $C_6 := \frac{1}{(q-p)C_5}$. (9) leads to $\Pr[MM < K] \sim e^{-e^{[ln\ n - K\ ln\ \mu_1 - ln\ C_6 + ln\ K]}}$.

Let $K = [ln\ n + ln\ ln\ n - ln\ C_6 - ln\ ln\ \mu_1 + \eta]/ln\ \mu_1$. (T5) is now immediate. Note that this corresponds to a periodic function of $[ln\ n + ln\ ln\ n]/ln\ \mu_1$. This kind of behaviour is rather common in data structures: see for instance Louchard and Szpankowski [10].

6 Conclusion

Using classical properties of RW and BM, we have derived asymptotic properties of a RW generation algorithm. Generalization to more than 3 steps is quite evident only with more algebra. We intend to pursue this approach on other walks generation algorithms. A preliminary version of this paper was presented at GASCOM 96.

Acknowledgements

The pertinent comments of the referee led to substantial improvements of the presentation.

References

[1] D. ALDOUS; *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York (1989).

[2] E. BARCUCCI, R. PINZANI, AND R. SPRUGNOLI; The Random generation of underdiagonal walks, *Discr. Math.*, **139** 3-18 (1995).

[3] E. BARCUCCI, R. PINZANI, AND R. SPRUGNOLI; Génération aléatoire des animaux dirigés, in *J. Labelle and J.-G. Penaud, eds., Actes de l'Atelier Franco-Québecois de Combinatoire, Publications du LACIM*, **Vol. 10** (1992) 17-25 (an upgraded English version will appear in Theoretical Computer Science).

[4] P. BILLINGSLEY; *Convergence of Probability Measures*, Wiley, New York, (1968).

[5] K.L. CHUNG; Excursions in Brownian motion, *Ark. Mat.*, **14** , 155-177 (1976).

[6] W. FELLER; *Introduction to Probability Theory and its applications*, Wiley, Vol. II, New York, (1970).

[7] K. ITO AND H.P. jr MCKEAN; *Diffusion Processes and their Sample Paths*; 2nd edn., Springer Verlag, Berlin (1974).

[8] G. LOUCHARD AND R. SCHOTT; Probabilistic Analysis of some Distributed Algorithms, *Random Structures and Algorithms*, **Vol. 2**, No 2 (1991).

[9] G. LOUCHARD; Kac's Formula, Levy's Local Time and Brownian Excursion, *J. Appl. Prob.*, **21**, 479-499 (1984).

[10] G. LOUCHARD AND W. SZPANKOWSKI; Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm, *IEEE Trans. Inf. Theor.*, **41**, 478-488 (1995).

[11] G. LOUCHARD; Some distributed algorithms revisited, *Stoch. Models*, **11 (4)**, 563-586 (1995).

[12] F. SPITZER, *Principles of Random Walk*, D. Van Nostrand Company, Inc., New York, (1964).