Probabilistic Analysis of a Schröder Walk Generation Algorithm

Guy Louchard * Olivier Roques †

May 2, 2000

Abstract

Using some tools from Combinatorics, Probability Theory, and Singularity analysis, we present a complete asymptotic probabilistic analysis of the cost of a Schröder walk generation algorithm proposed by Penaud et al.([13]). Such a walk $S(.)$ is made of northeast, southeast and east steps, but each east step is made of two time units (if we consider recording the time t on the abscissa and the moves on the ordinates). The walk starts from the origin at time 0, cannot go under the time axis, and we add the constraint $S(2n) = 0$. Five different probability distributions will appear in the study: Gaussian, Exponential, Geometric, Rayleigh and a new probability distribution, that we can characterize by its density Laplace Transform and its moments.

1 Introduction

In [11], we have analyzed some asymptotic properties of an underdiagonal walk generation algorithm (GA). This can be represented as a walk on a plane with northeast, southeast and east steps, under the condition that the walk cannot go under the x axis.

As announced in [11], we intend to pursue this approach on other walks GA. In this paper, we analyze an algorithm proposed in [13] to generate Schröder walks of length $2n$. Such a walk $S(.)$ is also made of northeast, southeast and east steps, but each east step is made of two time units (if we consider recording the time t on the abscissa and the moves on the ordinates). As in [11], the walk starts from the origin at time 0, cannot go under the time axis, but now, we add the constraint $S(2n) = 0$. We denote by R_n the number of S of length $2n$.

Let us say for completeness that $\{R_n\}_{n\geq 0}$ was discovered by Schröder ([14]) looking at the generalized bracketing problem. Several combinatorial objects are enumerated by R_n, such as : the number of dissections of a convex polygon, planted rooted trees with $n+1$ leaves whose nodes have a degree greater than one (sometimes called ordered hierarchies ([7]) in classification theory) and some string edit problems.

Another walk will be used in the sequel: $SL(.)$ is a left factor of a S i.e. a walk starting from the origin at time 0 and only constrained to remain above the time axis. We denote by F_n the number of SL of length n.

The GA proposed in [13] proceeds as follows. Using a slightly different form of the cyclic lemma ([4]) and the Catalan factorization ([8]), the authors establish a mapping between a $S(.)$ walk of length $2n$, and $(n+1)$ $SL(.)$ walks of length $2n+1$ which end either by a southeast step or by an east step. Hence, sampling uniformly in this subset of SL leads to a Schröder walk with the uniform distribution. The first part of the GA given in [13], that we will call GA.1, uses a rejecting method, which builds a walk step by step using some constant probability, such that the probability of a northeast step (resp. southeast step) is p (resp. p) and the probability of an east step is p^2, where p is the positive root

*Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, email: louchard@ulb.ac.be

†LaBRI, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence Cedex, France, email: roques@labri.u-bordeaux.fr

of the polynomial $1-2p-p^2$. GA.1 is allowed to fail (and then has to restart) either if the walk crosses the time axis, or if the walk stops at time $2n+2$ or if it stops at time $2n+1$ by a forbidden step (actually a northeast step). When GA.1 succeeds, it outputs a valid $SL(.)$ walk of height h, defined as the difference between the number of northeast steps and the number of southeast steps . Then, in the last part of GA (GA.2), a procedure called *lightning procedure* in [13], based on Catalan factorization, maps $\frac{h+1}{2}$ northeast steps into $\frac{h+1}{2}$ southeast steps, in order to make the walk ending at height -1. Finally, another combinatorial mapping, based on the cyclic lemma, maps the walk obtained previously in a Schröder walk.

The generating functions (GF) of R_n and F_n are given in [13] , respectively as

$$R(z) = \frac{1-z-\sqrt{1-6z+z^2}}{2z}, \tag{1}$$

$$F(z) = \frac{2}{1-2z-z^2+\sqrt{1-6z^2+z^4}}. \tag{2}$$

It is proved in [13] that the complexity of the proposed GA (in terms of number of used letters) is $\mathcal{O}(n)$. The purpose of this paper is to give a complete asymptotic ($n \to \infty$) precise probabilistic analysis of this GA (in terms of used letters and calls to a random generator).

In the field of analysis of algorithms, the moments of 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.

Using some tools from Combinatorics, Probability Theory, and Singularity analysis, we obtain a surprising list of 5 different probability distributions : Gaussian, Exponential, Geometric, Rayleigh and a new probability distribution, that we can characterize by its density Laplace Transform and its moments.

Another useful GF: $G(w,z)$ is related to the number Nt of transformed steps during the lightning procedure: marking these steps by w, we obtain (the proof is given in Sec 3.4), setting $f_1 := \sqrt{1-6z^2+z^4}$,

$$G(w,z) = \frac{2z(1-z^2+f_1+2wz^2)}{(-1+z^2-f_1+2\sqrt{w}\,z)(-1+z^2-f_1-2\sqrt{w}\,z)}. \tag{3}$$

The paper is organized as follows: Sec.2 presents the probabilistic aspects and the main results, Sec.3 gives some asymptotic analysis of several parameters and RV. Sec.4 is devoted to analysis of the GA. Sec.5 develops some simulation results and Sec.6 concludes the paper. An appendix provides rather technical proofs.

The following probabilistic notations will be used in the sequel:

- $\stackrel{\mathcal{D}}{\sim}$: convergence in distribution, for $n \to \infty$.
- $\Longrightarrow_{n\to\infty}$: weak convergence in the space of all right– continuous functions having left limits in R^2 and endowed with the Skorohod metric d_0 (see Billingsley, [2],Chapter III).
- $\mathcal{N}(M,V)$:= the Normal (or Gaussian) RV with mean M and variance V.
- Brownian Motion (BM) := Markovian Gaussian process, with mean $E[BM(t)] = 0$, variance $\mathrm{VAR}[BM(t)] = t$, and covariance $E[BM(s)\cdot BM(t)] = s (s \le t)$: see Ito, McKean, [10].
- GEOM(P) : geometric RV with probability distribution $P(1-P)^{k-1}$.
- Rayleigh : RV with density $e^{-x^2/2}x dx$.
- Exponential : RV with density e^{-x}.

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

2 Probabilistic aspects and main results

In this section we present the connections between S and some random walks (RW) and stochastic processes. We also summarize our main results.

The GA is based on the fact that we can associate to each S a RW such that all S with the same lenght are endowed with the same probability. This amounts to define three time-space steps (up, down and horizontal) with probabilities:

$$\begin{aligned} p &:= Pr[X_i = 1, \Delta t_i = 1] = Pr[X_i = -1, \Delta t_i = 1], \\ r = p^2 &:= Pr[X_i = 0, \Delta t_i = 2], \\ \text{and } p^2 + 2p &= 1, \text{ i.e. } p = \sqrt{2} - 1, r = 3 - 2\sqrt{2}. \end{aligned} \tag{4}$$

$S(i)$ is actually an Excursion, starting from 0, returning to 0 at time $2n$ and staying strictly positive for $0 < i < 2n$. $SL(i)$ is called a Meander, starting from 0 , conditionned to stay strictly positive for $0 < i \leq n$.

Let us mention a third RW, the Bridge $B(.)$ (with the same time-space steps) which has been used in [12], in connection with some string edit problems. There the only condition is that $B(n) = 0$. The probabilites (4) have also been derived in [12]. It is also proved in the same paper that the number Nh of horizontal steps in a Bridge such that $B(n) = \alpha\sqrt{n}, \alpha = \mathcal{O}(1)$, is characterized by

$$\frac{Nh - \mu_h \cdot n}{\sigma_h \sqrt{n}} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1), \quad n \to \infty, \quad \text{with} \tag{5}$$

$$\mu_h = (2 - \sqrt{2})/4, \quad \sigma_h^2 = \sqrt{2}/16. \tag{6}$$

Let us note that the asymptotic behaviour of Nh is *independent of* α. This is *not the case* for the number of up or down steps in the Bridge. Another result from [12] is related to the unconditioned RW $Y(.)$, endowed with (4) :

$$\frac{2^{1/4} Y([nt])}{\sqrt{n}} \Longrightarrow BM(t), \quad n \to \infty, \quad t \subset [0,1]. \tag{7}$$

We now present the main results of this paper. The following RV notations will be used in the sequel:

- κ_1 := number of letters necessary to generate a Schröder walk, where we associate a letter to each time unit.
- κ_2 := number of calls to a random generator necessary to generate a Schröder walk.
- T := number of trials in the rejecting procedure.
- Nh := number of horizontal steps in an Excursion or a Meander of length ℓ.
- Nr := number of Excursions generated before getting a length $2n$ Meander.
- Nt := number of transformed steps in the ligthning procedure.
- h := normalized(by $1/\sqrt{2n}$) height of the length $2n$ Meander.

Our main results are summarized in Table 1. We give the asymptotic distribution of the principal RV we have considered in the paper.

3 Some asymptotic analysis

In this section, we present a precise asymptotic analysis of several parameters and RV we need in the sequel. We analyze successively the Excursion, the Meander, $R(z)$ and $F(z)$, and the lightning procedure. A good introduction to the kind of techniques we need here can be found in Flajolet, Sedgewick [6].

RV	Asymptotic characteristics$(n \to \infty)$
κ_1	$\Psi_1(\alpha) := E[\exp(-\alpha\kappa_1/2n)] \sim \frac{1}{2}\phi_1(\alpha)/[1-1/2\phi_1(\alpha)]$ (ϕ_1 is given in (27))
$\kappa_2 \overset{\mathcal{D}}{\equiv} \mu_s\kappa_1$	$\Psi_2(\alpha) := E[\exp(-\alpha\kappa_2/2n)] = \Psi_1(\mu_s\alpha)$,(μ_s is given in (16))
T	$T \overset{\mathcal{D}}{\sim}$ GEOM(1/2)
Nh	$\frac{Nh-\ell\mu_h}{\sqrt{\ell}\sigma_h} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1), \ell \to \infty$ (μ_h, σ_h are given in (6))
Nr	$\frac{Nr \cdot 2^{1/4}}{\sqrt{\pi n}} \overset{\mathcal{D}}{\sim}$ Exponential RV
$Nt/\sqrt{2n} \overset{\mathcal{D}}{\equiv} h/2$	$Nt/\sqrt{2n} \overset{\mathcal{D}}{\sim}$ Rayleigh RV with density $exp(-x^2 2\sqrt{2})4x\sqrt{2}$

Table 1: Asymptotic characteristics $(n \to \infty)$.

3.1 Excursions

The hitting time to 0, T_0, starting from 1, has a probability generating function (PGF) $F_1(z)$, which satisfies the equation:

$$\begin{aligned} F_1 &= pz + pzF_1^2 + rz^2F_1, \quad \text{hence} \\ F_1(z) &= (1 - p^2z^2 - (1-6p^2z^2+p^4z^4)^{1/2}/(2pz). \end{aligned}$$

It is simpler to substitute $z^2 = u$ in $F_1(z) \cdot z = \varphi_1(u)$, say , and we have $Pr_1[T_0 = 2n-1] = [u^n]\varphi_1(u)$. Similarly, the PGF of $Pr_1[T_0 \geq 2n-1]$ is given by $\varphi_2(u) = \frac{1-\varphi_1(u)}{1-u}$.

The dominant algebraic singularity of φ_2 is given by 1. So we set $u = 1-\varepsilon$ and expand φ_2 into ε. This gives a Puiseux series:

$$\varphi_2(u) \sim 2^{1/4}/\sqrt{\varepsilon} - (\sqrt{2}-1)/2 + \mathcal{O}(\sqrt{\varepsilon}), \quad \varepsilon \to 0.$$

By classical singularity analysis, we obtain immediately

$$Pr_1[T_0 \geq 2n-1] = [u^n]\varphi_2(u) \sim \frac{2^{1/4}}{\sqrt{\pi n}}, \quad n \to \infty. \tag{8}$$

T_0 is of course directly related to the length of an Excursion.

Similarly, starting from $\varphi_1(u)$, we derive: $\varphi_1(u) \sim 1 - 2^{1/4}\sqrt{\varepsilon} + \mathcal{O}(\sqrt{\varepsilon})$ and

$$Pr_1[T_0 = 2n-1] \sim \frac{2^{1/4}}{2\sqrt{\pi}n^{3/2}}, \quad n \to \infty, \tag{9}$$

which is of course compatible with (8).

A more interesting RV is the number Nh of horizontal steps in an Excursion of length $2n$ (we mark Nh with w). The PGF $F_2(w,z)$ satisfies

$$\begin{aligned} F_2 &= pz + pzF_2^2 + rz^2wF_2, \quad \text{hence} \\ F_2(w,z) &= [1 - p^2z^2w - (1 - 2p^2z^2w + p^4z^4w^2 - 4p^2z^2)^{1/2}]/(2pz). \end{aligned}$$

Substitute again $z^2 = u$ in $F_2(w,z) \cdot z = \varphi_3(w,u)$, say. The dominant singularity is now given by

$$r(w) = (8\sqrt{2} + 12 + 6w + 4w\sqrt{2}) - 4(17 + 12\sqrt{2} + 12w\sqrt{2} + 17w)^{1/2}/(2w^2) \tag{10}$$

and $r(1) = 1$. More precisely

$$r(w) = 1 + C_1(w-1) + C_2(w-1)^2 + C_3(w-1)^3 + \mathcal{O}((w-1)^4), \quad w \to 1,$$

$$\begin{aligned} C_1 &= \sqrt{2}/2 - 1, \\ C_2 &= 5/4 - 13\sqrt{2}/16, \\ C_3 &= 69\sqrt{2}/64 - 25/16. \end{aligned} \tag{11}$$

Setting now $u = r(w) - \varepsilon$, we derive

$$\varphi_3(w,u) \sim G_1(w) + G_2(w)\sqrt{\varepsilon} + \mathcal{O}(\varepsilon), \quad \varepsilon \to 0,$$

with

$$\begin{aligned} G_1(w) &:= [2\sqrt{2} - 2 + (2 - 3\sqrt{2}/2)(w-1) + \mathcal{O}((w-1)^2)]/(2p) \sim 1 + \mathcal{O}(w-1), \quad w \to 1, \\ G_2(w) &:= -[2(3\sqrt{2}-4)^{1/2} + (3\sqrt{2}-4)^{1/2}/4(w-1) + \mathcal{O}((w-1)^2)]/(2p) \\ &\sim -2^{1/4} + \mathcal{O}(w-1), \quad w \to 1, \end{aligned} \tag{12}$$

and

$$[u^n]\varphi_3(w,u) \sim \frac{G_2(w)}{-2\sqrt{\pi n^3}} [1/r(w)]^n \sqrt{r(w)}, \quad n \to \infty.$$

Proceeding now as in Bender's Theorems 1 and 3 in [1], we set $w = e^{is}$ and expand $ln(1/r(w))$. This leads to

$$is(1 - \sqrt{2}/2) - \sqrt{2}/16 s^2 + \mathcal{O}(s^3), \quad s \to 0, \tag{13}$$

and finally, we obtain, normalizing by (9):

$$\frac{Nh - n(1 - \sqrt{2}/2)}{\sqrt{n\sqrt{2}/8}} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1), \quad n \to \infty, \tag{14}$$

which is identical to (5) , *even with different constaints.* The *total number* M of steps in an excursion of length ℓ is given by $M = \tilde{N} + Nh$, where $\tilde{N}$ is the number of up and down steps. Hence $\tilde{N} + 2Nh = \ell$ and , by (14),

$$\frac{Nh - \ell\mu_h}{\sqrt{\ell}\sigma_h} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1), \quad \ell \to \infty, \tag{15}$$

(μ_h, σ_h are given in (6)). Therefore,

$$\frac{M - l\mu_s}{\sqrt{l}\sigma_h} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1), \quad \ell \to \infty, \quad \text{where } \mu_s = (1/2 + \sqrt{2}/4). \tag{16}$$

It is finally clear from (8), that the number Nr of returns to 0 before getting a length $2n$ Meander (i.e. the number of Excursions) is such that

$$\frac{Nr \cdot 2^{1/4}}{\sqrt{\pi n}} \overset{\mathcal{D}}{\sim} \text{Exponential}, \quad n \to \infty, \tag{17}$$

with Probability Distribution Function (PDF) $1 - e^{-x}$. (Indeed, GEOM has an asymptotic distribution given by the Exponential RV). Actually, this corresponds to the local time of $Y(.)$ at the origin.

3.2 Meanders

Let us now turn to the length $2n$ Meander distribution. Reverting the time axis, this amounts to analyze the hitting time $T_0^{(h)}$ to 0 , starting from an height $h\sqrt{2n}$. The PGF is of course given by $[F_1(z)]^{h\sqrt{2n}}$ (this is equivalent to the ladder epochs: see Feller[5], Chap.XII.7).

Setting again $z^2 = u$ and $u = 1-\varepsilon$, we derive the PGF of $\tilde{T_0}^{(h)} = T_0^{(h)} + h\sqrt{2n}$ as

$$(\varphi_1(u))^{h\sqrt{2n}} \sim (1-2^{1/4}\sqrt{\varepsilon})^{h\sqrt{2n}}, \quad \varepsilon \to 0.$$

Now we set $u = e^{-\theta/n}$. We obtain asymptotically

$$e^{-\sqrt{\theta}h2^{1/4}\sqrt{2}}, \tag{18}$$

which means that $\tilde{T_0}^{(h)}/(2n)$ has, asymptotically, the density(wrt t):

$$h2^{1/4}exp(-h^2\sqrt{2}/(2t))/\sqrt{2\pi t^3}.$$

(We note that this is also true for $T_0^{(h)}/(2n)$). Setting $t=1$ and normalizing, this leads to the asymptotic (normalized by $1/\sqrt{2n}$) Meander (of duration $2n$) density, given by

$$e^{-h^2\sqrt{2}/2}h\sqrt{2}. \tag{19}$$

This is the well known Rayleigh density, which gives the classical BM Meander density. So (19) could have been derived from (7). The GF approach will however be necessary in the next analysis. The moments of h are given by

$$E[h^k] = 2^{k/4}\Gamma(k/2+1). \tag{20}$$

Note that the Rayleigh distribution appears in other asymptotic analysis: for instance the number of trees in a random mapping and the number of points at distance d to a cycle are also asymptotically Rayleigh: see Drmota and Soria [3].

Let us now consider the number Nh of horizontal steps in a Meander of length $2n$. Again we derive (The proof is given in Appendix A) the distribution (14). The number M of steps is again characterized by (16). In Appendix A we prove also the asymptotic independence of Nh and h.

3.3 R_n and F_n

The dominant algebraic singularity of $R(z)$ (see (1)) is given by $z_1^* = 3-2\sqrt{2} \equiv p^2$. Standard analysis gives, with $z = z_1^* - \varepsilon$,

$$R(z) \sim \sqrt{2}+1-2^{1/4}(3+2\sqrt{2})\sqrt{\varepsilon}+(17/2+6\sqrt{2})\varepsilon+2^{1/4}(-195\sqrt{2}-276)\varepsilon^{3/2}/16+\mathcal{O}(\varepsilon^2), \quad \varepsilon \to 0$$

Hence

$$R_n \sim \frac{C_4}{n^{3/2}z_1^{*n}} + \frac{C_5}{n^{5/2}z_1^{*n}}, \quad n \to \infty, \tag{21}$$

with

$$C_4 = 2^{1/4}(\sqrt{2}+1)/(2\sqrt{\pi}), C_5 = -3\cdot 2^{1/4}(14+11\sqrt{2})/(64\sqrt{\pi}).$$

We use the fact that

$$[z^n](1-z)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}(1+\frac{\alpha(\alpha-1)}{2n}+\mathcal{O}(\frac{1}{n^2})), \quad n \to \infty.$$

The dominant singularity of $F(z)$ is given by the root (with smallest module) of the denominator, i.e. $z_2^* = \sqrt{2}-1 \equiv p$, and, with $z = z_2^* - \varepsilon$,

$$F(z) \sim \frac{1}{[(4-2\sqrt{2})^{1/2}\sqrt{\varepsilon}]} - \sqrt{2}/(4-2\sqrt{2}) + \mathcal{O}(\sqrt{\varepsilon}), \quad \varepsilon \to 0.$$

Hence

$$F_n \sim \frac{C_4}{\sqrt{n} z_2^{*n}}, \quad n \to \infty. \tag{22}$$

To get more precise information on F_{2n} and F_{2n-1}, we extract from (2) their GF, given respectively by

$$\begin{aligned} F^+(z) &:= \frac{F(z)+F(-z)}{2} = \frac{2(1-z^2+f_1)}{(1-z^2+f_1)^2-4z^2}, \\ \text{with } f_1 &:= \sqrt{1-6z^2+z^4}, \\ F^-(z) &:= \frac{F(z)-F(-z)}{2} = \frac{4z}{(1-z^2+f_1)^2-4z^2}. \end{aligned}$$

Set again $z^2 = u$ and $u = z_1^* - \varepsilon$, this leads to

$$F^+ \sim 2^{3/4}/(4\sqrt{\varepsilon}) - 2^{1/4}/32\sqrt{\varepsilon} + \mathcal{O}(\varepsilon^{3/2}), \quad \varepsilon \to 0.$$

Hence

$$F_{2n} \sim \frac{1}{z_1^{*n}}[2^{1/4}(\sqrt{2}+2)/(4\sqrt{\pi n}) - (5+\sqrt{2})2^{1/4}/(64\sqrt{\pi}n^{3/2}) + \mathcal{O}(n^{-5/2})], \quad n \to \infty. \tag{23}$$

Also

$$zF^- \sim 2^{1/4}(2-\sqrt{2})/(4\sqrt{\varepsilon}) + 2^{1/4}(3\sqrt{2}+1)/32\sqrt{\varepsilon} + \mathcal{O}(\varepsilon^{3/2}), \quad \varepsilon \to 0.$$

Hence

$$F_{2n-1} \sim \frac{1}{z_1^{*n}}[2^{3/4}/(4\sqrt{\pi n}) - 5 \cdot 2^{1/4}/(64\sqrt{\pi}n^{3/2}) + \mathcal{O}(n^{-5/2})], \quad n \to \infty.$$

It is combinatorically proved in [13] that $(n+1)R_n = F_{2n} + F_{2n-1}$. This amounts to prove that $F^+(z) + [zF^-](z) = R(u) + uR'(u)$, which is easily checked. It is easy to check that the first two terms of $F_{2n} + F_{2n-1}$ coincide with the first two terms of $(n+1)R_n$.

3.4 Lightning Procedure

First of all let us prove (3), the GF of Nt. Let h be the height of the walk sampled by the first part GA.1 of the GA given in [13], which is always odd, because the length of the walk is $2n+1$. A northeast step $\{(i,j),(i+1,j+1)\}$ of the walk is said to be lighted iff there is no southeast step $\{(m,j+1),(m,j)\}$ in the walk, for $m > i+1$. The lightning procedure simply transforms each lighted step whose height is less or equal than $\frac{h+1}{2}$, into a southeast step. Note that there are exactly $\frac{h+1}{2}$ lighted steps of height less or equal than $\frac{h+1}{2}$ because lighted steps are precisely those which increase the height of the walk. Hence in order to establish (3), we consider the bivariate generating function $F(w,z)$ of $SL(.)$ walks, where z counts the length and w counts the height. From the functional equation for $F(z)$, given in [13], we have

$$F(w,z) = 1 + z^2R(z^2)F(w,z) + wzF(w,z) + z^2F(w,z),$$

where $R(z^2)$ is the generating function of $S(.)$ walks according to their length. The walk given by GA.1 is either a walk of length $2n$ in which a southeast step is added, or a walk of length $2n-1$ in which a (two times units) east step is added. Hence, the bivariate GF for this walk is

$$T(w,z) = \frac{z}{w}\frac{F(w,z)+F(w,-z)}{2} + z^2\frac{F(w,z)-F(w,-z)}{2}.$$

Now, if we want w to count only the half-height ($\frac{h+1}{2}$), we obtain

$$G(w,z) = \sqrt{w}\,T(\sqrt{w},z) = \frac{2z(1-z^2+f_1+2wz^2)}{(-1+z^2-f_1+2\sqrt{w}\,z)(-1+z^2-f_1-2\sqrt{w}\,z)},$$

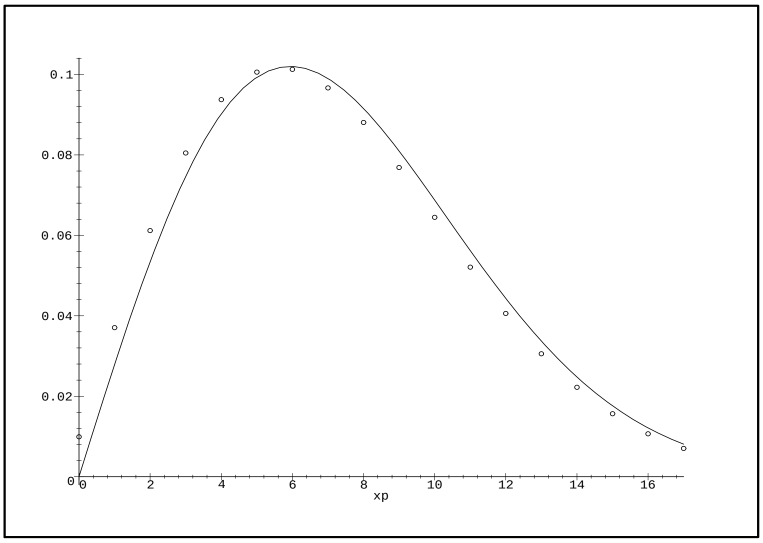

Figure 1: Observed and asymptotic normalized distribution of Nt

as announced before.
Substitute $z^2 = u$ in $G(w,z)/z = \varphi_4(w,u)$, say. Let us first note that, for $w = 1$, we recover the GF of $(n+1)R_n$. Also, if $w = 1$, the dominant singularity of $\varphi_4(1,u)$ is given by the root(with smallest module)of the denominator, i.e. $u^* = p^2 = 3 - 2\sqrt{2}$. The corresponding singularity of $\varphi_4(w,u)$ is given by

$$r(w) = [1 + 4w + w^2 - (1 + 8w + 14w^2 + 8w^3 + w^4)^{1/2}]/(2w)$$

and $r(1) = u^*$. More precisely,

$$r(w) = u^* + C_6(w-1)^2 + C_7(w-1)^3 + \mathcal{O}((w-1)^4), \quad w \to 1, \quad \text{with}$$
$$C_6 = (1/2 - 3/8\sqrt{2}), C_7 = (-1/2 + 3/8\sqrt{2}).$$

As usual, we set $u = r(w) - \varepsilon$. We obtain

$$\begin{aligned}
\varphi_4 &\sim \varphi_5/(\varphi_6\varepsilon) + \mathcal{O}(1), \quad \varepsilon \to 0, \quad \text{with} \\
\varphi_5 &= (w-1)[4 - 2\sqrt{2} + (5 - 3\sqrt{2})(w-1) + \mathcal{O}((w-1)^2)], \quad w \to 1, \\
\varphi_6 &= 4\sqrt{2} + 6\sqrt{2}(w-1) + \mathcal{O}((w-1)^2), \quad w \to 1,
\end{aligned}$$

and

$$[u^n]\varphi_4(w,u) \sim \varphi_5/[r(w)\varphi_6\ r(w)^n], \quad n \to \infty. \tag{24}$$

Set now $w = e^{i\xi/\sqrt{2n}}$ in (24). Normalizing by $(n+1)R_n$, this leads to the asymptotic equivalent

$$i\sqrt{2\pi}\xi e^{-\xi^2\sqrt{2}/16}(\sqrt{2}/8)^{1/2}.$$

This corresponds exactly to the Rayleigh density (19), with a 1/2 factor, i.e. $Nt/\sqrt{2n} \overset{\mathcal{D}}{\sim} h/2$, as expected.

Let us check numerically our results. For $n = 100$, Fig.1 gives the asymptotic and observed normalized distribution of Nt (observed=circle, asymptotic=line). The observed one is extracted from $[u^{100}]\varphi_4$. The asymptotic one is given by $\exp(-x^2 2\sqrt{2})4x\sqrt{2}$.

There is an obvious bias: this can be explained as follows. Proceeding as in Hwang [9] ,Theorem 2, we analyze $\varphi_7(w)/\varphi_7(1)$, where φ_7 is extracted from (24): $\varphi_7(w) = \varphi_5/(s\ r(w)\varphi_6)$. Set now $w = e^{is}$ in $\ln[\varphi_7(w)/\varphi_7(1)]$ and expand into s. This gives the dominant term $\frac{-\sqrt{2}}{4}is = -.3535539059381\ldots is = \beta \cdot is$, say and β is exactly the bias we have to reintroduce in Fig.1 ($n = 100$). This gives Fig.2, where the adjustment is now nearly perfect.

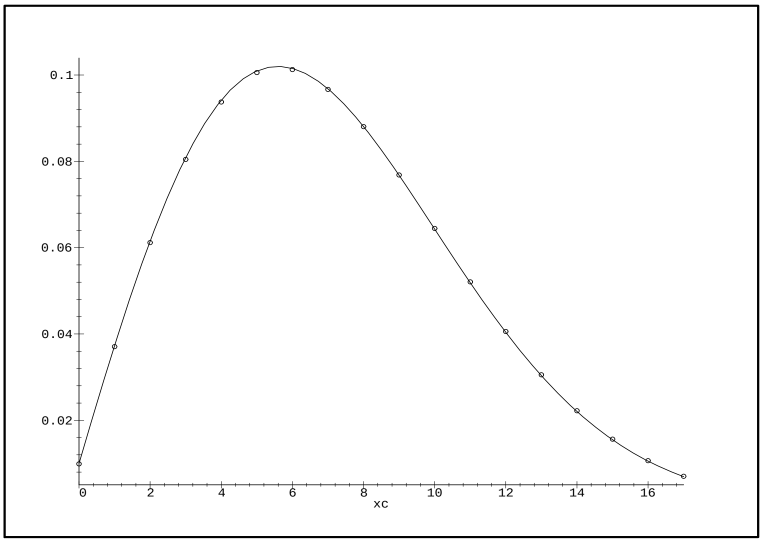

Figure 2: Observed and asymptotic normalized distribution of Nt, with bias

4 Asymptotic Analysis of the Generation Algorithm

In this section, we provide a complete asymptotic probabilistic analysis of the GA. We consider successively the rejecting procedure, the Meander generation, the Algorithm total cost and the cost moments.

4.1 Rejecting procedure

From the first part of the algorithm, we see that we accept the Meander with probability $P = A_1/(A_1 + A_2 + A_3)$, where

$$\begin{aligned} A_1 &:= (n+1)R_n p^{2n+1}, \\ A_2 &:= F_{2n} p^{2n+2}, \\ A_3 &:= F_{2n} p^{2n+1}. \end{aligned} \tag{25}$$

Hence $P \sim \frac{1}{2}, n \to \infty$, by (21) and (22). An interesting alternative proof of this result is given in Appendix B, based on a renewal argument.

The number T of trials to obtain a suitable walk is such that, asymptotically,

$$T \overset{\mathcal{D}}{\sim} \mathrm{GEOM(P)} \tag{26}$$

By (21) and (23), we obtain a more precise equivalent: $P \sim 1/2 + (4 - 3\sqrt{2})/(32n) + \mathcal{O}(1/n^2)$.

4.2 Meander Generation

In terms of *the number H_1 of generated letters*, the RV $\tilde{H}_1 := H_1/(2n)$ has been analyzed in [11], $T2$: this is the cost of generating a Meander of length $2n$. We have obtained the Laplace transform

$$\phi_1(\alpha) = E[\exp(-\alpha\tilde{H}_1)] \sim e^{-\alpha}/[e^{-\alpha} + \sqrt{\pi\alpha}\,\mathrm{erf}(\sqrt{\alpha})], \quad n \to \infty. \tag{27}$$

(This was established for a Motzkin path, but the analysis is similar). This corresponds to a length 1 Brownian Meander generation. In terms of *the number H_2 of calls to the random generator* we must use the *total* number of steps M in an Excursion or a Meander. But from (16), we see that only $E(M)$ must be taken into account: the random (Gaussian) part is $\mathcal{O}(\sqrt{n})$ and is asymptotically negligible after normalizing by $1/(2n)$. Hence we obtain, with $\tilde{H}_2 := H_2/(2n) : \tilde{H}_2 \overset{\mathcal{D}}{\equiv} \mu_s \tilde{H}_1$ and

$$\phi_2(\alpha) = E[\exp(-\alpha \tilde{H}_2)] \sim e^{-\alpha\mu_s}/[e^{-\alpha\mu_s} + \sqrt{\pi\alpha\mu_s}\,\mathrm{erf}(\sqrt{\alpha\mu_s})], \quad n \to \infty. \tag{28}$$

4.3 Total Cost

The total cost is made of two parts: the first part $\kappa_.$ is related to the Meander generation : its asymptotic normalized density Laplace transform, namely $E[exp(-\alpha\tilde{\kappa}_.)]$, where $\tilde{\kappa}_. := \kappa_./(2n)$, is given by

$$\Psi_.(\alpha) = P\phi_.(\alpha)/[1-(1-P)\phi_.(\alpha)], \tag{29}$$

where $\phi_.(\alpha) = \phi_1$ or ϕ_2, according to the cost elementary unit: letter or step.

The second part Nt is related to the ligthning procedure. The asymptotic normalized density of $Nt/\sqrt{2n}$ is given by $exp(-x^2 2\sqrt{2})4x\sqrt{2}$. As $Nt/\sqrt{2n} \overset{\mathcal{D}}{\equiv} h/2$, the moments of Nt are immediately derived from (20). The two costs are asymptotically independent.

4.4 Moments

The moments of Ψ_1 or Ψ_2 are easily computed. For instance, under Ψ_1 , with P:=1/2, $m_1 := E(\tilde{\kappa}_1) = 4, \mathrm{VAR}(\tilde{\kappa}_1) = 32/3$ and the third and fourth centered moments are given by $\mu_3(\tilde{\kappa}_1) = 352/5, \mu_4(\tilde{\kappa}_1) = 108544/105$. We just expand $e^{\alpha m_1}\Psi_.(\alpha)$ into α .

As a check, we extract from (21) the mean number of trials during the rejecting technique: this is given by

$$1/[(n+1)R_n p^{2n+1}] \sim \frac{\sqrt{n}}{C_4 p}, \quad n \to \infty. \tag{30}$$

The total mean number of letters used in a Meander generation is given by $E_1 + E_2 + E_3 + E_4$, with $E_1 := (2n+1)A_1, E_2 := (2n+2)A_2, E_3 := (2n+1)A_3$,

$$E_4 := p + \sum_{i=1}^{n-1}(2i+1)R_i p^{2i+1} \sim 2C_4 p \sum_{i=1}^{n-1}\frac{1}{\sqrt{i}} \sim 2C_4 p 2\sqrt{n}, \quad n \to \infty.$$

With (25), we obtain

$$E_6 := E_1 + E_2 + E_3 + E_4 \sim 2nC_4 2p/\sqrt{n} + 4C_4 p\sqrt{n} = 8C_4 p\sqrt{n}, \quad n \to \infty.$$

Multiplying by (30) leads to $8n = 2n.m_1$ as it should.

The moments of $\tilde{\kappa}_2$ are given by (as $\tilde{\kappa}_2 \overset{\mathcal{D}}{\equiv} \mu_s \tilde{\kappa}_1$)
$m_1 = 2+\sqrt{2}, \mathrm{VAR} = 4 + 8\sqrt{2}/3, \mu_3 = 22 + 77\sqrt{2}/5, \mu_4 = \frac{28832}{105} + \frac{6784}{35}\sqrt{2}$.

5 Simulations

We have done an extensive set of simulations, in order to check the quality of our asymptotic equivalents. For $n = 2000$, we have simulated $N = 1000$ generations. This has led to $NET = 134.553$ Excursions, $NM = 2083$ Meanders. To get reasonably long Excursions, we have observed in detail $NE = 11061$ Excursions of length ≥ 100.

The first picture Fig.3 gives (for $n = 30$) the RW $S(i)$. 4 Meanders were generated.

Fig.4 shows the Distribution Function (DF) of Nr : the number of returns to 0 (i.e. Excursions) before getting a length n=2000 Meander. According to (17), the limiting distribution is Exponential. The observations are based on NM generated Meanders.

The DF of the Meander height h is given in Fig.5, based on NM generated Meanders. According to (19), the limiting distribution is Rayleigh.

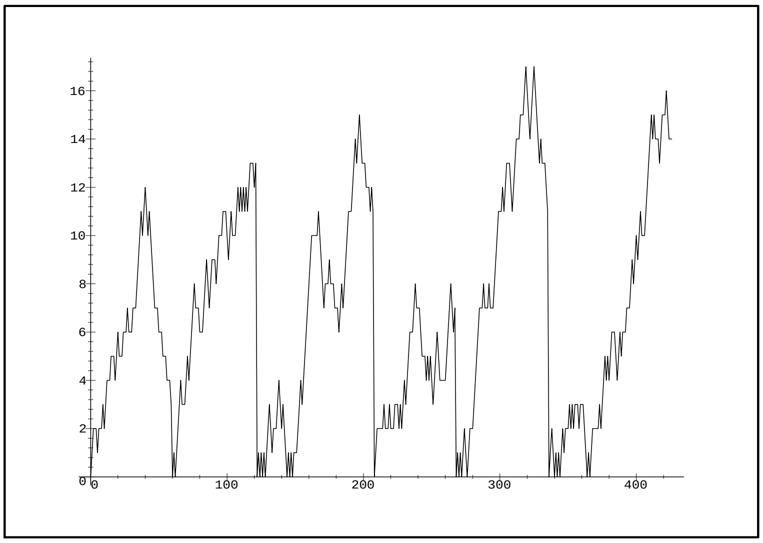

Figure 3: $S(i), n = 30$.

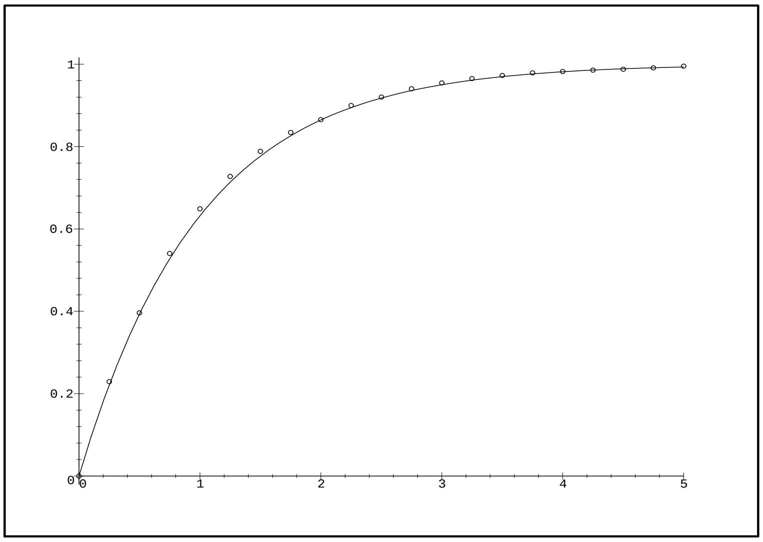

Figure 4: Nr : observed and limiting (exponential) DF.

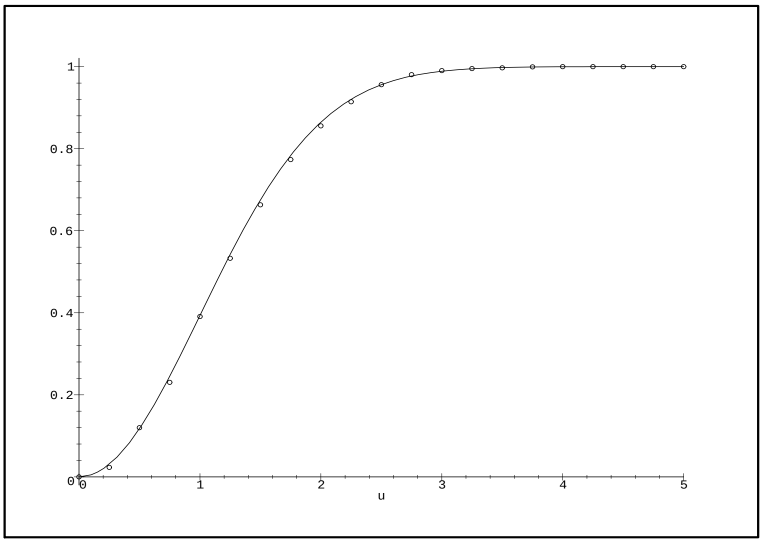

Figure 5: h : Observed and limiting (Rayleigh) DF.

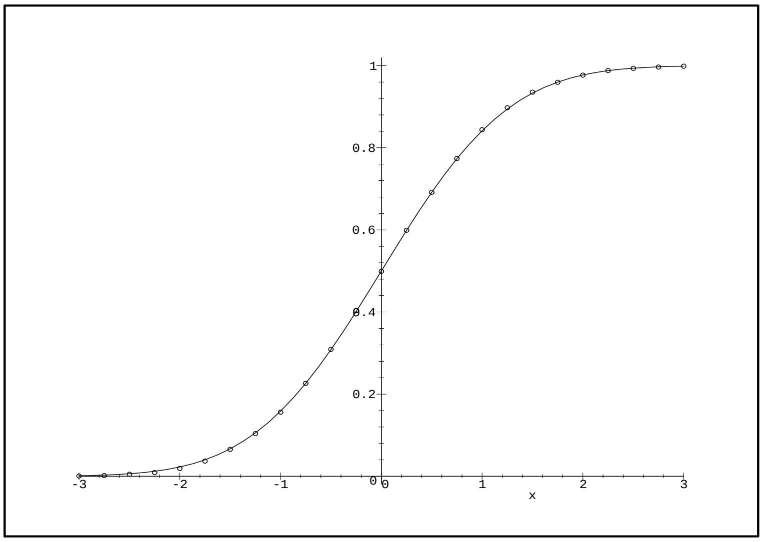

Figure 6: Nh in Excursions : observed and limiting (Gaussian) DF.

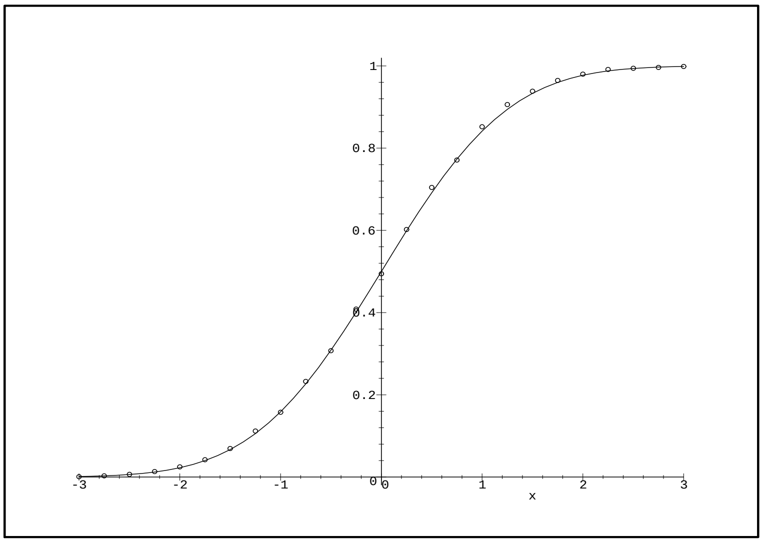

Figure 7: Nh in Meanders : observed and limiting (Gaussian) DF.

In Fig.6, we display the DF of Nh: the number of horizontal steps in Excursions, based on the NE Excursions of length ≥ 100. According to (15), the limiting distribution is Gaussian.

We have also checked in Fig.7 the same DF for the NM Meanders. Again we obtain a limiting Gaussian DF. Fig.8 shows the DF of T: the number of trials before getting a suitable Meander, based on N generations. According to (26), we obtain a GEOM(1/2) distribution. The fit is not as good as in our previous pictures. The observed frequency at 1 is .472. The corresponding standard deviation $\sigma = \sqrt{1/(4N)} = 0.016\cdots$ So the observed frequency falls within 1.77σ of $1/2$, which is statistically acceptable.

In Fig.9, we give the observed and limiting Laplace transform $\Psi_1(\alpha)$ of the Meander generation cost (according to the number of generated letters). The fit with the expression given in (29) is quite good.

Fig.10 shows the same observation for the steps cost (number of calls to the random generator). Here we consider $\Psi_2(\alpha)$.

6 Conclusion

Using various tools from Combinatorics, Probability Theory, and Singularity analysis, we have achieved a complete asymptotic probabilistic analysis of the cost of a Schröder walk generation algorithm proposed by Penaud et al. ([13]). Five different probability distributions have been observed. We intend to pursue such approach in other generation algorithms found in the litterature.

References

[1] E. Bender. Central and local limit theorems applied to asymptotics enumeration. *Journal of Combinatorial Theory,* Series A, 15:91–111, 1973.

[2] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[3] M. Drmota and M. Soria. Marking in combinatorial constructions: Generating functions and limiting distributions. *Theoretical Computer Science*, 144(1–2):67–100, 1995.

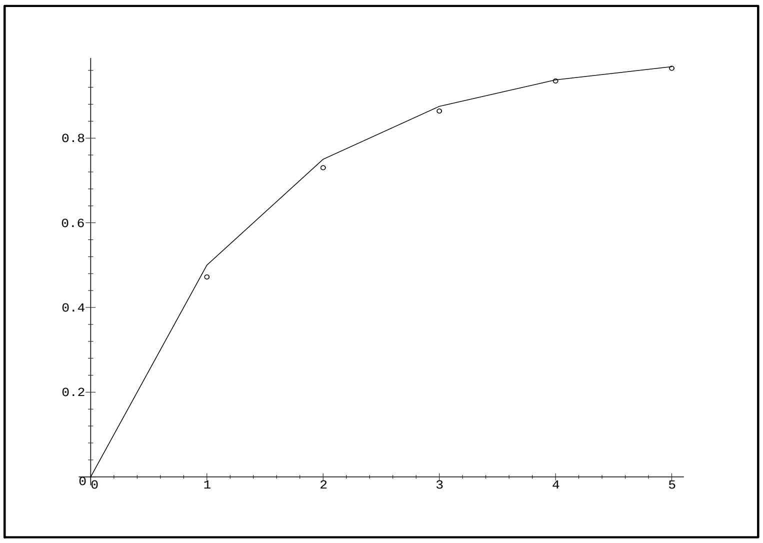

Figure 8: T : observed and limiting (GEOM(1/2)) DF.

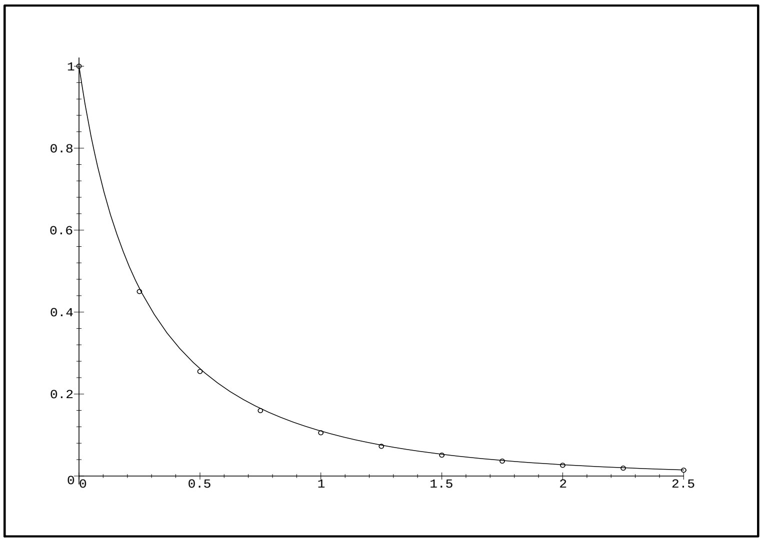

Figure 9: Cost(letters) : observed and limiting ($\Psi_1(\alpha)$) Laplace transform.

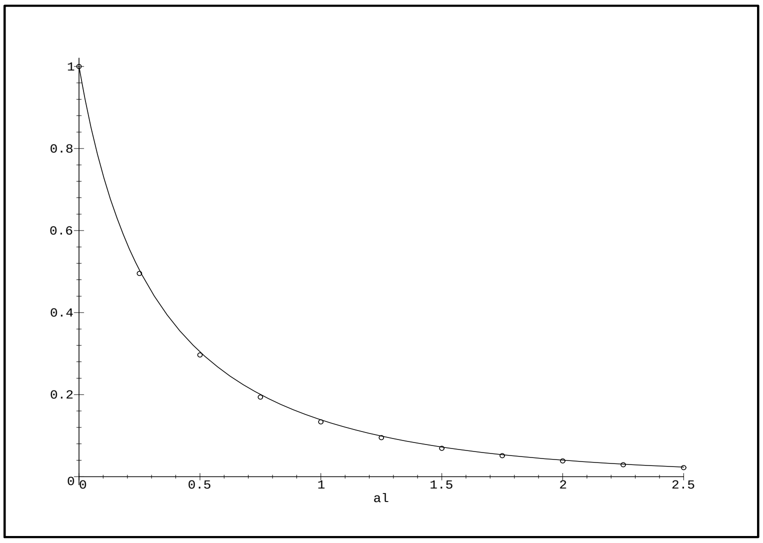

Figure 10: Cost(steps) : observed and limiting ($\Psi_2(\alpha)$) Laplace transform.

[4] A. Dvoretzky and T. Motzkin. A problem of arrangements. *Duke Mathematical Journal*, 14:305–313, 1947.

[5] W. Feller. *Introduction to Probability Theory and its Applications.* Wiley, 1970.

[6] P. Flajolet and R. Sedgewick. The average case analysis of algorithms: Multivariate asymptotics and limit distributions. *INRIA T.R.*, 3162, 1997.

[7] P. Flajolet, P. Zimmerman, and B. V. Cutsem. A calculus for the random generation of labelled combinatorial structures. *Theoretical Computer Science*, 132:1–35, 1994.

[8] D. Gouyou-Beauchamps and G. Viennot. Equivalence of the two-dimensional directed animal problem to a one-dimensional path problem. *Advances in Applied Mathematics*, 9:334–357, 1988.

[9] H. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, To appear.

[10] K. Ito and H. McKean, Jr. *Diffusion Processes and their Sample Paths.* Springer, 1974.

[11] G. Louchard. Asymptotic properties of some underdiagonal walks generation algorithms. *Theoretical Computer Science*, 218:249–262, 1999.

[12] G. Louchard and W. Szpankowski. A probabilistic analysis of a string edit problem and its variations. *Combinatorics, Probability and Computing*, 4:143–166, 1995.

[13] J. Penaud, E. Pergola, R. Pinzani, and O. Roques. Chemins de Schröder et hiérarchies aléatoires. *Theoretical Computer Science*, To appear.

[14] E. Schröder. Vier combinatorische probleme. *Z. Math. Phys.*, 15:361–370, 1870.

A Appendix: Proof of (14) for Meanders.

Proof Set $\tau := T_0^{(h)}/(2n)$. We must compute

$$[u^\beta][\varphi_3(w,u)]^{h\sqrt{2n}}, \tag{31}$$

where β is such that $n \cdot \tau = \beta - h\sqrt{2n}/2$. Now we analyze

$$\begin{aligned} \varphi_3(w,u) &:= (1-p^2uw)/(2p) - [G_3(w,u)]^{1/2}/(2p), \text{with} \\ G_3(w,u) &:= 1-2p^2uw+p^4u^2w^2-4p^2u. \end{aligned} \tag{32}$$

First, we set $w = e^{is}, u = e^{-\theta}$ in the first part of (32). This leads to $G_4(s) + \theta G_5(s) + \mathcal{O}(\theta^2)$, with $G_4(s) = 1 - isp/2, G_5(s) = p/2 + \mathcal{O}(s)$.
We should be tempted to proceed immediately in the same way for G_3. But this is not correct: as we shall see, we need a more precise analysis, based on the GF. So we set $u = 1 - \varepsilon$, this gives

$$G_3 = [(-11+8\sqrt{2}) + (4\sqrt{2}-6)w + (17-12\sqrt{2})w^2] + [(12-8\sqrt{2}) + (6-4\sqrt{2})w + (-34+24\sqrt{2})w^2]\varepsilon$$

$$+\mathcal{O}(\varepsilon^2), \quad \varepsilon \to 0.$$

But, if $\varphi(u) = \sum p_k u^k$, then $[u^k][\varphi(u\alpha)] = p_k\alpha^k$. If we want to cancel the constant term $G_3(w,1)$ in G_3, we must choose α such that $G_3(w,\alpha) = 0$, namely $\alpha = r(w)$ as given by (10). This leads to

$$\tilde{G}_3 := G_3(w, r(w)u) = \varepsilon G_6(w) + \mathcal{O}(\varepsilon^2), \quad \varepsilon \to 0,$$

with

$$G_6 = -2p^4 r(w)^2 w^2 + 2p^2 r(w) w + 4p^2 r(w).$$

Now we can set $w = e^{is}, u = e^{-\theta}$. This leads to

$$\tilde{G}_3 = \theta[-16 + 12\sqrt{2} + is(24 - 17\sqrt{2}) + \mathcal{O}(s^2)] + \mathcal{O}(\theta^2), \quad \theta \to 0,$$

and

$$\sqrt{\tilde{G}_3}/(2p) = (2^{1/4} + \mathcal{O}(s))\sqrt{\theta} + \mathcal{O}(\theta^{3/2}), \quad \theta \to 0.$$

Now setting $u = r(w)u$ amounts to replace θ in the first part of (32) by $\theta - [is(\sqrt{2}/2 - 1) + \mathcal{O}(s^2)]$, by (13). Therefore, we obtain, for the probability distribution of τ, the asymptotic expression

$$\left\{[G_4(s) + (\theta - is(\sqrt{2}/2-1) + \mathcal{O}(s^2))G_5(s) - (2^{1/4} + \mathcal{O}(s))\sqrt{\theta} + \mathcal{O}(\theta^{3/2})]^{h\sqrt{2n}}\right\}^{(-1)} e^{-\beta\ ln[r(w)]}, \quad \theta \to 0,$$

where $\varphi^{(-1)}(\theta)$ is the probability density corresponding to $\varphi(\theta)$. This leads to

$$\begin{aligned} &\Big\{\exp[h\sqrt{2n}[-isp/2 - is(\sqrt{2}/2-1)p/2 + \mathcal{O}(s^2) - (2^{1/4} + \mathcal{O}(s))\sqrt{\theta} + \mathcal{O}(\theta)] \\ &\quad -(n\tau + h\sqrt{2n}/2)[is(\sqrt{2}/2-1) + \sqrt{2}/16s^2 + \mathcal{O}(s^3)]\Big\}^{(-1)}, \quad \theta \to 0. \end{aligned}$$

First of all, the $\sqrt{n}s$ term disappears(as it should). Setting now $\theta = \theta/n$, we obtain, asymptotically, for the θ contribution:

$$e^{-\sqrt{\theta}h2^{1/4}\sqrt{2}},$$

which is identical to (18). We are left with

$$\exp[n\tau[is(1-\sqrt{2}/2) - \sqrt{2}/16s^2] + \mathcal{O}(ns^3) + \mathcal{O}(\sqrt{n}s^2)], \quad s \to 0,$$

which proves (14). ■

We observe that, asymptotically, for fixed τ, Nh is conditionally independent of h. This can be explained probabilistically as follows. Let us return to the original RW $Y(.)$. From (5), in $Y(n)$, $Nh \stackrel{\mathcal{D}}{\sim} n\mu_h + \sigma_h\sqrt{n}\xi_1$, where $\xi_1 = \mathcal{N}(0,1)$. The number of up and down steps

$Nv = n - 2Nh \sim n(1-2\mu_h) - 2\sigma_h\sqrt{n}\xi_1$ and $Y(n) \sim \sum_1^{Nv} V_i$, with $V_i = \pm 1$, with probability 1/2. Hence $Y(n) \sim \sqrt{Nv}\xi_2$, with $\xi_2 = \mathcal{N}(0,1)$ (independently of ξ_1). So

$$
\begin{aligned}
E(e^{i[xY(n)+yNh]}) &\sim E(e^{-Nv\ x^2/2+iyNh}) \\
&\sim E\{\exp[-x^2/2(n\sqrt{2}/2 - 2\sigma_h\sqrt{n}\xi_1) + iy(n\mu_h + \sigma_h\sqrt{n}\xi_1)]\} \\
&\sim \exp[-x^2/2n\sqrt{2}/2 + iyn\mu_h - \sigma_h^2 n/2[x^2/i + y]^2] \\
&= \exp(iyn\mu_h)\cdot\exp[-x^2/2n\sqrt{2}/2 - n\sigma_h^2/2[x^2/i+y]^2].
\end{aligned}
$$

Now we replace x by $x/\sqrt{n}$ and y by $y/\sqrt{n}$ in the second term. We are left, asymptotically, with $\exp[-x^2/2\sqrt{2}/2 - \sigma_h^2 y^2/2]$, which shows the asymptotic independence of $Y(.)$ and Nh (incidentally, we recover (7)).

B Appendix: A Renewal Approach

If we look only at the time displacement D, we have $Pr[D=1] = 2p, Pr[D=2] = p^2$. The renewal theorem (for discrete RV) gives the following DF for the residual waiting time (see Feller [5],XI.4): $H(j) = \sum_0^{j-1}[1-F(k)]/m_1$, where F is the DF of D and m_1 is its mean. We obtain $H(0) = 0$, $H(1) = 1/m_1$, $H(2) = 1$. Now let Ph be the stationary distribution of an horizontal step covering time n ($n \to \infty$) and let $Pv := 1 - Ph$. We must have $H(1) = Ph.1 + Pv.2p$, hence $Ph = (2-\sqrt{2})/4$, $Pv = (2+\sqrt{2})/4$. The asymptotic probability P of success is given by $Ph + Pv \cdot p = 1/2$ as expected.