On the importance sampling of self-avoiding walks

Mireille Bousquet-Mélou

To cite this version:

HAL Id: hal-00599400
https://hal.archives-ouvertes.fr/hal-00599400v2
Submitted on 24 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
ON THE IMPORTANCE SAMPLING OF SELF-AVOIDING WALKS

MIREILLE BOUSQUET-MÉLOU

Abstract. In a 1976 paper published in Science, Knuth presented an algorithm to sample (non-uniform) self-avoiding walks crossing a square of side \( k \). From this sample, he constructed an estimator for the number of such walks. The quality of this estimator is directly related to the (relative) variance of a certain random variable \( X_k \). From his experiments, Knuth suspected that this variance was extremely large (so that the estimator would not be very efficient). But how large? For the analogous Rosenbluth algorithm, which samples unconfined self-avoiding walks of length \( n \), the variance of the corresponding estimator is believed to be exponential in \( n \).

A few years ago, Bassetti and Diaconis showed that, for a sampler à la Knuth that generates walks crossing a \( k \times k \) square and consisting of North and East steps, the relative variance is only \( O(\sqrt{k}) \). In this note we take one step further and show that, for walks consisting of North, South and East steps, the relative variance jumps to \( 2^k (k+1)/(k+1)^2 \). This is quasi-exponential in the average length of the walks, which is of order \( k^2 \). We also obtain partial results for general self-avoiding walks crossing a square, suggesting that the relative variance could be exponential in \( k^2 \) (which is again the average length of these walks).

Knuth’s algorithm is a basic example of a widely used technique called sequential importance sampling. The present paper, following Bassetti and Diaconis’ paper, is one of very few examples where the variance of the estimator can be found.

1. Introduction

A self-avoiding walk (SAW) on a graph is a walk that never visits the same vertex twice. Let \( \mathcal{W}_k \) be the set of SAWs on a \( k \times k \) square grid, going from the South-West vertex to the North-East vertex (Figure 2). In his paper “Coping with finiteness” [14, 15], Knuth described the following algorithm to generate a (non-uniform) random walk of \( \mathcal{W}_k \): start from the South-West corner, and at each time, choose with equal probability (which can be 1/3, 1/2 or 1) one of the eligible steps. A step is eligible if, once appended to the current walk, it gives a self-avoiding walk that can be extended so as to end at the North-East corner. In this way the walk is never trapped and the algorithm always succeeds\(^1\). Figure 1 shows the probabilities of the 12 possible walks when \( k = 2 \). Two bigger examples (\( k = 10, k = 100 \)) are shown in Figure 2. This procedure is a basic example of a widely used technique called sequential importance sampling [5, 10, 11]. It is also a variant, for walks confined to a square, of the Rosenbluth algorithm that generates unconfined SAWS [21].

Denote by \( p(w) \) be the probability to draw the walk \( w \in \mathcal{W}_k \). Consider the random variable \( X_k = 1/p(w) \), where \( w \) is a random walk of \( \mathcal{W}_k \) drawn according to the distribution \( p(\cdot) \). Clearly,

\[
\mathbb{E}(X_k) = \sum_{w \in \mathcal{W}_k} p(w) \frac{1}{p(w)} = |\mathcal{W}_k|.
\]

Hence one can estimate the number of SAWS crossing a \( k \times k \) square by generating \( N \) walks \( w^{(1)}, \ldots, w^{(N)} \) of \( \mathcal{W}_k \), and computing

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{p(w^{(i)})} \quad (1)
\]

\(^{1}\)We describe in Section 5.3 how to detect algorithmically when a new step traps the walk.
2,2,1,2 2,2,3,1 2,2,3,1
2,2,3,1,1,1 2,2,1,2,2,1 2,2,1,2,2,1,1,1

Figure 1. The 12 self-avoiding walks crossing the $2 \times 2$ square. For each of them, we give the sequence $1/p_1, 1/p_2, \ldots$ where $p_i$ is the probability of the $i$th step. The probability of the walk is thus the reciprocal of the product of the terms in the list. Two walks have probability $1/8$, six have probability $1/12$, and four have probability $1/16$. Two walks that differ by a diagonal symmetry have the same probability.

By generating “several thousand” walks for $k = 10$, Knuth obtained

$$|W_{10}| \simeq (1.6 \pm 0.3) \times 10^{24},$$

which is quite good compared to the now known exact value, $1, 568, 758, 030, 464, 750, 013, 214, 100$ (see [7, 15]). We have reproduced Knuth’s experiment, and found, with a first group of 10,000 walks, the estimate $1.78 \times 10^{24}$, and with a second group, the estimate $1.38 \times 10^{24}$. As observed by Knuth, the values $1/p(w(i))$ vary a lot (a small sample of 10 walks gave us values ranging from $10^{11}$ to $10^{24}$), and one may suspect that the variance of $X_k$ is probably much larger than $E(X_k)^2$, or, in other words, that the relative variance

$$\text{Var} \left( \frac{X_k}{E(X_k)} \right)$$

is large. Note that

$$\text{Var}(X_k) = E(X_k^2) - E(X_k)^2 = \sum_{w \in W_k} \frac{1}{p(w)} - |W_k|^2.$$
Also, observe that the variance of the estimator (1) is \( \Var(X_k) / N \), so that \( \Var(X_k) \) is a measure of the quality of this estimator. Let us mention that, even though sequential importance sampling is widely used, no general bounds on the variance of the estimators are available.

Knuth’s observation led Bassetti and Diaconis to study a simpler algorithm, in which a step, to be eligible, has to go North (N) or East (E) [5]. The resulting walk is called a directed walk, or NE-walk. Each step has probability 1/2 unless it follows the North or East side of the square — in which case it has probability 1. Denote by \( D_k \) the set of directed walks of \( W_k \), and by \( p(w) \) the probability to generate the directed walk \( w \) with this new algorithm. Define the random variable \( X_k = 1/p(w) \) as above. Then

\[
\mathbb{E}(X_k) = |D_k| = \binom{2k}{k} \sim \frac{4^k}{\sqrt{\pi k}}.
\]

Of course, since \( |D_k| \) is known exactly, there is no point in using importance sampling to estimate this cardinality — but it is interesting to know that the variance of the estimator can be determined, as follows. By the above argument, a walk of \( D_k \) that hits the North or East side of the square for the first time at time \( k + i \) has probability \( 1/2^{k+i} \). Since there are \( 2^{k+i-1} \) such walks,

\[
\mathbb{E}(X_k^2) = \sum_{w \in D_k} \frac{1}{p(w)} = \sum_{i=0}^{k-1} 2^{k+i+1} \binom{k+i-1}{i}.
\]

The corresponding generating function is

\[
\sum_{k \geq 1} \mathbb{E}(X_k^2) x^k = \frac{2x}{1+2x} \left( \frac{3}{\sqrt{1-16x}} - 1 \right),
\]

and an elementary singularity analysis [12, Chap. VII] gives

\[
\mathbb{E}(X_k^2) \sim \frac{16^k}{3\sqrt{\pi k}},
\]

which is roughly \( \sqrt{k} \) times larger than

\[
\mathbb{E}(X_k)^2 \sim \frac{16^k}{\pi k}.
\]

In this note, we first take one more step in the direction of the general problem by declaring that South steps are also eligible. The resulting walks are partially directed walks, or NES-walks. The probabilities of the 9 walks obtained when \( k = 2 \) are shown in Figure 3. Of course, these probabilities are not the same as those obtained from Knuth’s original algorithm.

![Figure 3](image-url)

**Figure 3.** The nine NES-walks crossing a 2 \( \times \) 2 square, with the reciprocals of the probabilities of their steps.
We will prove that one outcome of this increased generality is that the ratio between $E(X_k^2)$ and $E(X_k)^2$ becomes much larger:

$$E(X_k^2) \sim \frac{3}{2} 2^{k(k+1)} \quad \text{while} \quad E(X_k)^2 = (k + 1)^{2k}.$$  

We will also prove that the average length of a (uniform) NES-walk confined to the $k \times k$ square is quadratic in $k$, so that the variance of $X_k$ is roughly exponential in the length, as predicted for SAWs generated by the Rosenbluth algorithm [6].

Since the $x/y$ symmetry is lost with partially directed walks, it is natural to generalize the original question by enclosing walks in a rectangle $R$ of height $k$ and width $\ell$. Thus, let $P_{k,\ell}$ be the set of partially directed walks that start from the South-West corner of $R$ and end at the North-East corner. A walk of $P_{k,\ell}$ contains exactly $\ell$ East steps, and choosing the heights of these steps determines the walk completely. Hence the number of walks in $P_{k,\ell}$ is $(k + 1)^\ell$.

Thus, defining the random variable $X_{k,\ell} = 1/p(w)$ as above, we have

$$E(X_{k,\ell}) = |P_{k,\ell}| = (k + 1)^\ell. \quad (3)$$

We will prove that, if $k, \ell \to \infty$ in such a way $\ell = o(2^k)$, then

$$\text{Var}(X_{k,\ell}) \sim E(X_{k,\ell}^2) \sim \frac{3}{2} 2^{(k+1)\ell},$$

so that the relative variance satisfies

$$\text{Var} \left( \frac{X_{k,\ell}}{E(X_{k,\ell})} \right) \sim \frac{3}{2} \left( \frac{2^{k+1}}{(k + 1)^2} \right)^\ell.$$

This result thus extends the very short list of examples where the variance of an importance sampler can be rigorously established [5]. Moreover, the average length of a (uniform) walk of $P_{k,\ell}$ is shown to be approximately $k\ell/3$, so that the variance is again exponential in the length, as expected for other similar samplers.

The paper is organized as follows. In Section 2, we obtain an explicit expression for the generating function of the numbers $E(X_{k,\ell}^2)$ (the counterpart of (2)). In Section 3, we derive from this expression the above asymptotic result. In Section 4, we go back to Knuth’s sampler and prove that there exist two positive constants $\lambda$ and $\beta$ such that

$$E(X_k)^{1/k^2} \to \lambda \quad \text{and} \quad E(X_{k,\ell}^2)^{1/k^2} \to \beta.$$  

The former result has actually been known since 1978 [1]. Since a variance is non-negative, $\beta \geq \lambda^2$. Upper (and lower) bounds on $\lambda$ have been obtained in [7], based on the determination of the numbers $E(X_k) = |W_k|$ for small values of $k$, and of related numbers counting other configurations of self-avoiding walks. As shown in Section 4, a similar study, performed for the numbers $E(X_{k,\ell}^2)$, might suffice to prove that $\beta \geq \lambda^2$, so that the variance would be again exponential in $k^2$ (which is known to be the average length of a uniform SAW crossing the $k \times k$-square [16]). We conclude with a few remarks and questions on the importance sampling of self-avoiding walks not confined to a box.

2. Exact results for NES-walks

In this section, we first describe the probability $p(w)$ to obtain the walk $w$ in terms of the geometry of $w$ (Section 2.1). This description reduces the determination of the numbers $E(X_{k,\ell}^2)$ to the enumeration of NES-walks according to several parameters, which we perform in Section 2.2.

2.1. The probability $p(w)$

Let $w_0$ be a walk of $P_{k,\ell}$, written as a sequence of $N$, $E$ and $S$ steps. Let $w$ be the prefix of $w_0$ that precedes the last $E$ step. That is, $w_0 = wEN \cdots N$. By convention, $w_0$ starts at height 0.
Lemma 1. The probability of obtaining \( w_0 \) via the importance sampling algorithm satisfies
\[
\frac{1}{p(w_0)} = 2 \cdot 3^{h(w)} 2^{h_c(w)} \cdot v(w) \cdot v_c(w),
\]
where
- \( h(w) \) is the number of horizontal steps of \( w \) that lie neither at height 0 nor at height \( k \),
- \( h_c(w) \) is the number of horizontal contacts of \( w \), that is, horizontal steps that lie at height 0 or \( k \),
- \( v(w) \) is the number of vertical steps of \( w \) that end neither at height 0 nor at height \( k \),
- \( v_c(w) \) is the number of vertical contacts of \( w \), that is, vertical steps that end at height 0 or \( k \).

Proof. Assume the walk \( w_0 \) has length \( n \) and ends with exactly \( j \) vertical steps. The probability of the first step is 1/2, and the probability of each of the \( j \) final steps is 1. Let \( s_i \) denote the \( i \)th step. Hence \( w \) consists of the steps \( s_1, \ldots, s_{n-j-1} \). For \( 1 \leq i < n-j \), the probability of \( s_{i+1} \) depends on the direction and position of \( s_i \):
- if \( s_i \) is horizontal, but not a contact, then the probability of \( s_{i+1} \) is 1/3,
- if \( s_i \) is a horizontal contact, then the probability of \( s_{i+1} \) is 1/2;
- if \( s_i \) is vertical, but not a contact, then the probability of \( s_{i+1} \) is 1/2,
- if \( s_i \) is a vertical contact, then the probability of \( s_{i+1} \) is 1.

The lemma follows.

2.2. Enumeration of NES-walks in a Strip of Fixed Height

Recall the expression (3) of the numbers \( \mathbb{E}(X_{k,\ell}) \). For \( k \) (the height of the rectangle) fixed, the generating function of the numbers \( \mathbb{E}(X_{k,\ell}^2) \) is rational:
\[
\sum_{\ell \geq 1} \mathbb{E}(X_{k,\ell})^2 x^\ell = \sum_{\ell \geq 1} (k + 1)^2 x^\ell = \frac{(k + 1)^2 x}{1 - (k + 1)^2 x}.
\]

We will determine the variance of \( X_{k,\ell} \) by describing the generating function of the numbers \( \mathbb{E}(X_{k,\ell}^2) \), which is also rational when \( k \) is fixed.

Proposition 2. For any fixed height \( k \), the generating function \( M_k(x) \) of the numbers \( \mathbb{E}(X_{k,\ell}^2) \) is a rational series:
\[
M_k(x) := \sum_{\ell \geq 1} \mathbb{E}(X_{k,\ell}^2) x^\ell = 2x \frac{N_k}{G_k},
\]
where \( N_k \) and \( G_k \) are polynomials in \( x \) satisfying the same recurrence relation:
\[
N_k = (5 + 9x)N_{k-2} - 4N_{k-4},
\]
(\( N_k \) and \( G_k \) have \( n \) and an initial conditions
\[
N_1 = 2, \quad G_1 = 1 - 4x, \\
N_2 = 5 + 3x, \quad G_2 = 1 - 9x - 6x^2, \\
N_3 = 11 + 9x, \quad G_3 = 1 - 19x - 18x^2, \\
N_4 = 23 + 54x + 27x^2, \quad G_4 = 1 - 36x - 99x^2 - 54x^3.
\]

Example. For \( k = 2 \),
\[
\sum_{\ell \geq 1} \mathbb{E}(X_{2,\ell}^2) x^\ell = 2x \frac{5 + 3x}{1 - 9x - 6x^2} = 10x + 96x^2 + O(x^3).
\]
Figure 3 allows us to check that the coefficient of \( x^2 \) is correct:
\[
96 = 4 + 8 + 8 + 12 + 12 + 8 + 12 + 16 + 16.
\]
The generating function of the variances is

\[ \sum_{\ell \geq 1} \text{Var}(X_{2,\ell})x^\ell = 2x \frac{5 + 3x}{1 - 9x - 6x^2} - \frac{9x}{1 - 9x}. \]

Observe that the radius of the first fraction is smaller than the radius of the second fraction. As \( \ell \to \infty \),

\[ E(X^2_{2,\ell}) \sim \mu^\ell \]

(up to a multiplicative constant) with \( \mu = (9 + \sqrt{105})/2 \simeq 9.62 \), while \( E(X_{2,\ell})^2 = 9^\ell \).

We now want to prove Proposition 2. Recall that

\[ E(X^2_{k,\ell}) = \sum_{w_0 \in P_{k,\ell}} \frac{1}{p(w_0)}. \]

The expression of \( p(w_0) \) given in Lemma 1 leads us to study a purely enumerative problem. For \( k \) fixed, let \( \mathcal{T}_k \) be the set of NES-walks \( w \) that start at height 0 and are confined to the strip \( 0 \leq y \leq k \). We wish to count these walks by the parameters \( h(w), h_c(w), v(w) \) and \( v_c(w) \). So, let

\[ \sum_{w \in \mathcal{T}_k} x^{h(w)}y^{v(w)}a^{h_c(w)}b^{v_c(w)} \]

be the associated generating function. This series is easily seen to be rational (it can be determined using a transfer-matrix approach [22, Sec. 4.7], or equivalently a finite-state automaton [12, p. 362]), and there are several ways to determine it. We present here what we believe to be the most direct one. It relies on a recursive description of the walks of \( \mathcal{T}_k \), where we add at each time an \( E \) step and a sequence of vertical steps. This approach requires to take into account an additional parameter, namely the height \( f(w) \) of the final point of the walk \( w \). Hence our series finally involve 5 variables:

\[ T_k(s) = T_k(x, y, a, b; s) = \sum_{w \in \mathcal{T}_k} x^{h(w)}y^{v(w)}a^{h_c(w)}b^{v_c(w)}s^{f(w)}. \]

We will denote by \( \tilde{T}_k \) the subset of \( \mathcal{T}_k \) formed of walks that do not end at height 0 or \( k \), and by \( \tilde{T}_k(s) \equiv \tilde{T}_k(x, y, a, b; s) \) the corresponding generating function. Accordingly,

\[ T_k(s) = \sum_{i=0}^{k} T_{k,i}s^i = T_{k,0} + \tilde{T}_k(s) + s^k T_{k,k}, \]

where \( T_{k,i} \) is the series in \( x, y, a \) and \( b \) counting walks of \( \mathcal{T}_k \) ending at height \( i \). By Lemma 1,

\[ M_k(x) = \sum_{\ell \geq 1} \mathbb{E}(X^2_{k,\ell})x^\ell = \sum_{\ell \geq 1} \sum_{w_0 \in P_{k,\ell}} \frac{1}{p(w_0)}x^\ell = 2\sum_{w \in \mathcal{T}_k} \sum_{\ell \geq 1} 3^{h(w)}2^{h_c(w)}2^{v_c(w)}x^\ell = 2x T_k(3x, 2, 2x, 1, 1). \]

Remark. The series \( T_{k,k} \) has already been determined in the case \( x = y = a = b = t \), using the same approach as here [3, Prop 3]. The derivation is more involved here because we keep track of four parameters in the enumeration, and because we are interested in \( T_k(k) \) rather than \( T_{k,k} \).

Lemma 3. The series \( \tilde{T}_k(s), T_{k,0} \) and \( T_{k,k} \) satisfy the following system of equations:

\[ \left(1 - \frac{x}{1 - ys} - \frac{xy^k}{1 - y^k}\right) \tilde{T}_k(s) = \frac{ys - (ys)^k}{1 - y^k} \tilde{T}_k(1/y) - \frac{x}{1 - ys} \tilde{T}_k(y) + a T_{k,0} \frac{ys - (ys)^k}{1 - y^k} + a T_{k,k} \frac{y^{k-1} - y^k}{1 - y^k}, \]
\[ T_{k,0} = 1 + bx \tilde{T}_k(y) + aT_{k,0} + aby^{k-1}T_{k,k}, \]
\[ T_{k,k} = by^{k-1} + bx y^{k-1} \tilde{T}_k(1/y) + aby^{k-1}T_{k,0} + aT_{k,k}, \]
with \( \bar{s} = 1/s \) and \( \bar{y} = 1/y \).

**Proof.** We construct the walks of \( \tilde{T}_k \) recursively, by adding at each time a horizontal step followed by a sequence of vertical steps.

We partition the set \( \tilde{T}_k \) into three disjoint subsets, illustrated in Figure 4.

- The first subset consists of walks with no \( E \) step. These walks consist of \( i \) North steps, with \( 1 \leq i < k \). Their generating function is
  \[ \sum_{i=1}^{k-1} (ys)^i = ys - (ys)^k \frac{1}{1 - ys}. \]
  - The second subset consists of walks in which the last \( E \) is followed by a (possibly empty) sequence of \( N \) steps. We denote by \( i \) the height of the last \( E \) step, and distinguish the cases \( i = 0 \) and \( 0 < i < k \). The generating function of this subset of \( \tilde{T}_k \) reads
    \[ aT_{k,0} \sum_{j=1}^{k-1} (sy)^j + x \sum_{i=1}^{k-1} \left( T_{k,i} s^i \sum_{j=0}^{k-1-i} (ys)^j \right) = aT_{k,0} \frac{ys - (ys)^k}{1 - ys} + x \sum_{i=1}^{k-1} \left( T_{k,i} s^i \frac{1 - (ys)^{k-i}}{1 - ys} \right) \]
    \[ = aT_{k,0} \frac{ys - (ys)^k}{1 - ys} + \frac{x}{1 - ys} \left( \tilde{T}_k(s) - (ys)^k \tilde{T}_k(1/y) \right). \]
  - The third subset consists of walks in which the last \( E \) step is followed by a non-empty sequence of \( S \) steps. We denote by \( i \) the height of the last \( E \) step, and distinguish the cases \( i = k \) and \( 0 < i < k \). The generating function of this subset of \( \tilde{T}_k \) reads
    \[ a s^k T_{k,k} \sum_{j=1}^{k-1} (y\bar{s})^j + x \sum_{i=1}^{k-1} \left( T_{k,i} s^i \sum_{j=0}^{k-1-i} (y\bar{s})^j \right) = aT_{k,k} \frac{ys^{k-1} - y^k}{1 - y\bar{s}} + x \sum_{i=1}^{k-1} \left( T_{k,i} s^i \frac{y\bar{s} - (y\bar{s})^i}{1 - y\bar{s}} \right) \]
    \[ = aT_{k,k} \frac{ys^{k-1} - y^k}{1 - y\bar{s}} + \frac{x}{1 - y\bar{s}} \left( y\bar{s} \tilde{T}_k(s) - \tilde{T}_k(y) \right). \]

Adding the three contributions gives the series \( \tilde{T}_k(s) \) and establishes the first equation of the lemma.

The equations for \( T_{k,0} \) and \( T_{k,k} \) are obtained in a similar fashion.

---

**Figure 4.** Recursive construction of bounded NES-walks.

We now solve the functional equations of Lemma 3. The key tool is the *kernel method* (see e.g. [4, 8, 20]).

**Proposition 4.** Let \( k \geq 1 \). The series \( T_k(1) \equiv T_k(x, y, a; b; 1) \) counting NES-walks confined to a strip of height \( k \) is

\[ T_k(1) = \frac{N_k}{G_k}, \]
where \( N_k \) and \( G_k \) are polynomials in \( x, y, a \) and \( b \) satisfying the same recurrence relation
\[
N_k = (1 - x + y^2(1 + x))N_{k-2} - y^2N_{k-4},
\]
(and similarly for \( G_k \)) with initial conditions
\[
N_{-1} = (1 - x - xy)(b - y)/y^2, \quad G_0 = (x - 1)ab/y - (x + 1)(a - 1),
\]
\[
N_0 = (b - xb + xy)/y, \quad G_1 = 1 - a - ab,
\]
\[
N_1 = 1 + b, \quad G_2 = (1 - x)(1 - a) - (x + 1)yab,
\]
\[
N_2 = 1 - x + y + by(1 + x), \quad G_3 = (1 - x - xy)(1 - a) - yab(x + y + xy).
\]
Equivalently,
\[
T_k(1) = \frac{1}{P(S)S^k + P(S)} \left( Q(\bar{S})S^k + Q(S) + (1 - y^2)\frac{S^k - S}{S - 1} \right),
\]
where \( S \) is the unique formal power series\(^2\) in \( x \) and \( y \) satisfying
\[
S + \frac{1}{S} = (1 + x)y + (1 - x)\bar{y},
\]
with \( \bar{y} = 1/y, \bar{S} = 1/S \),
\[
P(s) = 1 - a + aby - s(ab + y - ay) \quad \text{and} \quad Q(s) = 1 - by + (b - y)s.
\]
The reason why we give the expressions of \( N_{-1} \) and \( N_0 \), rather than \( N_3 \) and \( N_4 \), is that they are more compact. The same reason explains why we give \( G_0 \) rather than \( G_4 \). It is of course easy to compute \( N_3, N_4 \) and \( G_4 \), and Proposition 2 then follows at once, using (5). We hope that using the same notation \( N_k, G_k \), for the enumeration problem of Proposition 4 and its specialization of Proposition 2 will not create any confusion.

**Proof.** First, we use the last two equations of Lemma 3 to express \( \tilde{T}_k(y) \) and \( \tilde{T}_k(1/y) \) as linear combinations of \( T_{k,0} \) and \( T_{k,k} \). Then, in the first equation of the lemma, we replace \( \tilde{T}_k(y) \) and \( \tilde{T}_k(1/y) \) by their expressions in terms of \( T_{k,0} \) and \( T_{k,k} \). The left-hand side is unchanged, and the right-hand side now involves only two unknown series, namely \( T_{k,0} \) and \( T_{k,k} \):
\[
(1 - \frac{x}{1 - y} - \frac{xy\bar{y}}{1 - \bar{y}^2}) \tilde{T}_k(s) = \frac{ys}{1 - y^2} + \frac{y}{b(1 - y)} + \left( \frac{ays}{1 - y^2} + \frac{y(a - 1)}{b(1 - y)} \right) T_{k,0} + s^k \left( \frac{y(a - 1)}{b(1 - y)} + \frac{ya\bar{y}}{1 - y^2} \right) T_{k,k}.
\]
The kernel of this equation is the coefficient of \( \tilde{T}_k(s) \). It vanishes when \( s = S \) and \( s = \bar{S} := 1/S \), where \( S \) is defined in the proposition. Since \( \tilde{T}_k(s) \) is a polynomial in \( s \), and \( S \) and \( \bar{S} \) are Laurent series in \( x \) and \( y \) with finitely many monomials with negative exponents, the series \( \tilde{T}_k(S) \) and \( \tilde{T}_k(\bar{S}) \) are well-defined. Replacing \( s \) by \( S \) or \( \bar{S} \) in the above equation cancels the left-hand side, and hence the right-hand side. One thus obtains two linear equations between \( T_{k,0} \) and \( T_{k,k} \), which involve the series \( S \). Solving them gives expressions of \( T_{k,0} \) and \( T_{k,k} \) in terms of \( S \) (the expression of \( T_{k,k} \) is given in (11) below). By setting \( s = 1 \) in (9), one then expresses \( \tilde{T}_k(1) \) in terms of \( S \), and finally \( T_k(1) = T_{k,0} + T_{k,k} + \tilde{T}_k(1) \). This gives (6).

Observe that the expression (6) is unchanged if we replace \( S \) by \( S = 1/S \). In particular, it can be written as a symmetric rational function in \( S \) and \( \bar{S} \) (with coefficients in \( Q(a, b, y) \)). Since \( S \) and \( \bar{S} \) are the two roots of (7), their symmetric functions are rational functions of \( x \) and \( y \). This implies that \( T_k(1) \) is a rational series in \( x, y, a \) and \( b \). However, the denominator of (6),

---

\(^2\)The other solution is \( 1/S \), and its expansion in \( x \) and \( y \) involves negative powers of \( y \).
namely $P(\bar{S})S^k + P(S)$, is not unchanged when $S \mapsto 1/S$. But let us define the series $N_k$ and $G_k$ as follows:

$$G_{2k} = \frac{y^k}{1-y^2} (P(\bar{S})S^k + P(S)\bar{S}^k),$$

$$G_{2k+1} = \frac{y^k}{(1-y)(1+S)} (P(S)S^{k+1} + P(S)\bar{S}^{k+1}),$$

$$N_{2k} = \frac{y^k}{1-y^2} \left( Q(S)S^k + Q(S)\bar{S}^k + (1-y^2)\frac{S^k - \bar{S}^{k-1}}{S-1} \right),$$

$$N_{2k+1} = \frac{y^k}{(1-y)(1+S)} \left( Q(S)S^{k+1} + Q(S)\bar{S}^{k+1} + (1-y^2)\frac{S^k - \bar{S}^k}{1-S} \right).$$

Then it is easy to check that (6) can be rewritten as $T_k(1) = N_k/G_k$. Moreover, the series $N_k$ and $G_k$ are unchanged when $S \mapsto 1/S$, and hence, by the same argument as above, they are rational functions of $x$, $y$, $a$ and $b$. More precisely, each of the sequences $G_{2k}$, $G_{2k+1}$, $N_{2k}$ and $N_{2k+1}$ is of the form $y^k(\alpha S^k + \beta \bar{S}^k)$, where $S$ and $\bar{S}$ are the two roots of (7). Hence each sequence satisfies the recurrence relation

$$u_k = (1 - x + y^2(1 + x))u_{k-1} - y^2 u_{k-2}.$$ 

One easily determines the initial values for each sequence. This yields the description of $N_k$ and $G_k$ given in the proposition. From this description, it is clear that $N_k$ and $G_k$ are polynomials, as soon as $k \geq 1$.

**Remarks**

1. **Denominators.** The series $T_{k,i}$, counting walks ending at height $i$, are also rational, but with a denominator that is a proper multiple of the denominator of $T_k(1) = \sum_{i=0}^{k} T_{k,i}$. For instance,

$$T_{k,k} = \frac{b(y^2 - 1)S^k - P(S)S^k}{(P(S)S^k + P(S))},$$

or, in terms of polynomials,

$$T_{k,k} = \frac{b y^{k-1}}{F_k},$$

where $F_k$ is defined by the recurrence relation

$$F_1 = (1 - a - ab)(1 - a + ab) \quad \text{and} \quad F_2 = (1 - a + by)(1 - (1-a) - (x+1)yab).$$

It is not hard to prove that $G_k$, the denominator of $T_k(1)$, is a divisor of $F_k$. The simplification that occurs in the denominator when summing the series $T_{k,i}$ over $i$ has recently been explained combinatorially, for slightly different walk models, by Bacher [2].

2. **Average length.** The series $T_{k,k}(x,t,x,t,x)$ counts NES-walks crossing a strip of height $k$ according to the length (variable $l$) and the width (variable $x$). In particular, the case $t = 1$ of (12) reads $T_{k,k}(x,1,1,1) = 1/(1 - (k+1)x)$, as justified combinatorially in the introduction. In order to determine the average length $|w|$ of a uniform NES-walk $w$ crossing a $k \times \ell$ rectangle, we differentiate $T_{k,k}(x,t,x,t,t)$ with respect to $t$, and then set $t = 1$. This gives

$$\frac{\partial}{\partial l} \left( T_{k,k}(x,t,x,t,t) \right)_{t=1} = \sum_{\ell \geq 0} x^{\ell} \sum_{w \in P_{k,\ell}} |w| = \frac{k}{1-x(1+k)} + \frac{x(1+k)(1+k(1+k+2)x/3)}{(1-x(1+k))^2},$$

so that the average length is

$$\frac{(k^2 + 5k + 3) \ell}{3(1+k)} + \frac{k(2k+1)}{3(1+k)} = \frac{k\ell}{3} + O(k+\ell).$$  

(13)
3. Asymptotic results for NES-walks

We now derive asymptotic results from the previous section. Recall that $\mathbb{E}(X_{k,\ell}) = (k + 1)^\ell$, so that the generating function of the numbers $\mathbb{E}(X_{k,\ell})^2$ is given by (4) (for $k$ fixed), with radius of convergence $1/(k + 1)^2$. The radius of convergence of the generating function of the numbers $\mathbb{E}(X_{k,\ell}^2)$ turns out to be exponentially smaller. Our study has analogies with the study of the longest run in a binary string [12, p. 308], which also requires to analyze (explicit) rational functions depending on an integer $k$.

Proposition 5. Let $k \geq 1$. The series $M_k(x)$, given in Proposition 2, has a unique pole $\rho_k$ of modulus less than $1/9$, satisfying

$$\rho_k = \frac{1}{2^{k+1}} + \frac{9}{2 \cdot 4^{k+1}} - 12k - 23 + \frac{36k^2 - 54k - 87/8}{16^{k+1}} + O \left( \frac{k^3}{32k} \right).$$

As $x \to \rho_k^-$,

$$M_k(x) \sim \frac{\alpha_k}{1 - x/\rho_k},$$

with

$$\alpha_k = \frac{3}{2} - \frac{9k - 4}{2k^2} + \frac{27k^2 - 48k + 1}{2 \cdot 4k^2} = 81k^3 - 306k^2 + 75k + 140 + O \left( \frac{k^4}{16k} \right).$$

The second moment of $X_{k,\ell}$ satisfies, uniformly in $k$ and $\ell$,

$$\mathbb{E}(X_{k,\ell}^2) = \alpha_k \rho_k^{-1} + O(9^\ell k).$$

In particular, if $k, \ell \to \infty$ in such a way $\ell = o(2^k)$, then

$$\text{Var}(X_{k,\ell}) \sim \mathbb{E}(X_{k,\ell}^2) \sim \frac{3}{2} 2^{(k+1)\ell},$$

which is much larger than $\mathbb{E}(X_{k,\ell})^2 = (k + 1)^{2\ell}$. By (13), the variance is thus exponential in the average length of a (uniform) NES-walk crossing the $k \times \ell$ rectangle.

Proof. We proceed in four steps. We first express the series $M_k(x)$ in terms of an algebraic series $S$, as was done for the enumerative problem in Proposition 4. Then, we study the analytic properties of $S$. We use these properties to prove that the denominator $G_k$ of $M_k$ is real-rooted, with one positive zero $\rho_k$ and all the other zeroes below $-1/9$. We finally apply Cauchy’s formula to extract the $\ell$th coefficient of $M_k(x)$, which is $\mathbb{E}(X_{k,\ell}^2)$.

Step 1. The expression of $M_k$

By Proposition 2,

$$M_k(x) = 2x \frac{N_k}{G_k},$$

where the polynomials $N_k$ and $G_k$ can be described either by induction, or, after performing the change of variables $x \to 3x$, $a \to 2x$, $y \to 2$ and $b \to 1$ in (10), by\(^3\)

$$G_{2k} = \begin{cases} \frac{2^k}{3} \left( P(1/S)S^k + P(S)S^{-k} \right), \\ \frac{2^k}{1 + S} \left( P(1/S)S^{k+1} + P(S)S^{-k} \right) \end{cases},$$

$$G_{2k+1} = \begin{cases} \frac{2^k}{3} \left( Q(1/S)S^k + Q(S)S^{-k} - 3\frac{S^k - S^{-k}}{1 - 1/S} \right), \\ \frac{2^k}{1 + S} \left( Q(1/S)S^{k+1} + Q(S)S^{-k} - 3\frac{S^k - S^{-k}}{1 - 1/S} \right) \end{cases},$$

\(^3\)From now on, we carefully avoid the notation $S := 1/S$, since we will soon be doing complex analysis.
where $S$ and $1/S$ are the two power series in $x$ satisfying
\[ S + \frac{1}{S} = \frac{5 + 9x}{2}, \]  
(18)
or equivalently,
\[ x = -\frac{1}{9}(2S - 1)(2/S - 1). \]  
(19)
The polynomials $P(s)$ and $Q(s)$ are
\[ P(s) = 1 + 2x - 2s(1 - x) \quad \text{and} \quad Q(s) = -1 - s, \]
so that, in view of (19),
\[ P(S) = \frac{(2S - 1)(2S^2 - 11S - 4)}{9S} \quad \text{and} \quad P(1/S) = \frac{(2 - S)(2 - 11S - 4S^2)}{9S^2}. \]
It also follows from (5) and (6) that
\[ M_k(x) = \frac{2x}{P(S)S^k + P(1/S)} \left( Q(S)S^k + Q(1/S) - 3 \frac{S^k - S}{S - 1} \right). \]  
(20)

**Step 2. The series $S(x)$**

From now on, we denote by $S$ the root of (18) that has constant term $1/2$:
\[ S = \frac{5 + 9x - 3\sqrt{(1 + x)(1 + 9x)}}{4}. \]  
(21)

**Lemma 6.** The series $S$ has radius of convergence $1/9$, and admits an analytic continuation, still denoted by $S$, in $\mathbb{C} \setminus [-1, -1/9]$. In this domain, $S$ never vanishes, and its modulus is less than $1$.

**Proof.** The existence of an analytic continuation follows from basic complex analysis. If $x = u + iv$, the imaginary part of the discriminant $(1 + x)(1 + 9x)$ reads $2v(5 + 9u)$. Using the principal determination of the square root, the analytic continuation of $S$ is given by (21) when $\Re(x) \geq -5/9$, and otherwise by
\[ S = \frac{5 + 9x + 3\sqrt{(1 + x)(1 + 9x)}}{4}. \]

A plot of the modulus of $S$ is shown in Figure 5 (left).

![Figure 5](image)

**Figure 5.** Left: A plot of the modulus of $S$, showing the cut on the interval $[-1, -1/9]$. Right: The function $R(s)$.

**Step 3. The roots of $G_k$**
Lemma 7. For $k \geq 1$, the denominator $G_k$ of the series $M_k$ is real-rooted. It has a unique positive zero $\rho_k$, which, as $k \to \infty$, admits the expansion (14). The other zeroes are smaller than $-1/9$. As $x$ approaches $\rho_k$, the series $M_k$ behaves like $\alpha_k/(1-x/\rho_k)$, where $\alpha_k$ admits the expansion (16).

We could use Rouché’s theorem to prove that, for any $\varepsilon > 0$, the polynomial $G_k$ has only one root of modulus less than $1/9 - \varepsilon$ for $k$ large enough, but the above statement is more precise.

Proof. The case $k = 1$ being trivial ($G_1 = 1-4x$), we focus on the case $k \geq 2$. By Proposition 2, the denominator $G_k$ has degree $[\frac{k+1}{2}]$. The expressions (17) are symmetric in $S$ and $1/S$, and thus hold for any determination of $S$, and thus for any $x \in \mathbb{C}$, including in the cut $[-1, -1/9]$. They show that $G_k(x) = 0$ if and only if $S \neq -1$ and

$$S^{k-1} = \frac{(2S - 1)(2S^2 - 11S - 4)}{(2 - S)(2 - 11S - 4S^2)}.$$ 

Conversely, if $s \in \mathbb{C} \setminus \{-1\}$ is a root of

$$s^{k-1} = \frac{(2s - 1)(2s^2 - 11s - 4)}{(2 - s)(2 - 11s - 4s^2)} := R(s), \quad (22)$$

then

$$x := -\frac{1}{9}(2s - 1)(2/s - 1) \quad (23)$$

is a root of $G_k$. Observe that in this case, $1/s$ is also a root of (22), and gives rise to the same root $x$ of $G_k$. Conversely, if two distinct roots $s_0$ and $s_1$ of (22) give rise to the same root of $G_k$, then $s_1 = 1/s_0$.

It is easy to relate the positions of $s$ and $x$ in the complex plane. By writing $s = u + iv$, one finds that $x$ is real if and only if $s$ is real or has modulus 1. If $s = e^{i\theta}$, then $x = (4\cos\theta - 5)/9$ lies in $[-1, -1/9]$. If $s$ is real and negative, then $x \leq -1$, and the equality holds if and only if $s = -1$. If $s$ is real and positive, then $x \geq -1/9$, and $x > 0$ if and only if $s \not\in [1/2, 2]$.

Since we want to prove that $G_k$ is real-rooted, let us study the roots of (22), distinct from $-1$, that are real or have modulus 1. We will prove that (22) has

- two pairs $\{s, 1/s\}$ of real zeroes distinct from $-1$, one positive outside of $[1/2, 2]$, and one negative,
- $[\frac{k-1}{2}]$ pairs of zeroes distinct from $-1$ on the unit circle.

Consequently, $G_k$ has two real zeroes outside the interval $[-1, -1/9]$, one positive, one less than $-1$, and $[\frac{k-1}{2}]$ zeroes in $[-1, -1/9]$. In particular, it is real rooted.

Real roots of (22). An elementary study of the function $R(s)$, for $s \in \mathbb{R}$, reveals that it consists of 4 decreasing branches, shown in Figure 5 (right), with vertical asymptotes at

$$s = -\frac{11 + 3\sqrt{17}}{8} \simeq -2.9, \quad s = -\frac{11 + 3\sqrt{17}}{8} \simeq 0.17, \quad \text{and} \quad s = 2.$$

The branches intersect the $s$-axis at the reciprocals of these three values (and in particular at $1/2$). Thus in $\mathbb{R}^+$, the equation $s^{k-1} = R(s)$ has two roots, one below $1/2$ and the other beyond $2$, which are necessarily the reciprocal of each other. The smallest of these increases to $1/2$ as $k$ increases; thus the corresponding value of $x$ decreases to 0 as $k$ increases. We denote by $\rho_k$ this root of $G_k$.

If $k$ is even, the equation $s^{k-1} = R(s)$ has also two roots in $\mathbb{R}^-$. If $k \geq 3$ is odd, the curve $s \mapsto s^{k-1}$ intersects the second branch of $R(s)$ at $s = -1$, but also somewhere between $s = 0$ and $s = -0.508 \ldots$ (which is the root obtained for $k = 3$). The latter intersection point gives rise to a root of $G_k$ smaller than $-1$.

The rest of the argument will show that all other roots of $G_k$ lie in $[-1, -1/9]$.
Roots of (22) of modulus 1. We first observe that, if $s$ has modulus 1, then the same holds for $R(s)$. More precisely, if $s = e^{i\theta}$, then $R(s) = e^{i\theta}$ with
\[
\cos \phi = -\frac{56 + 321 \cos \theta - 336 \cos^2 \theta + 128 \cos^3 \theta}{(5 - 4 \cos \theta)(157 + 44 \cos \theta - 32 \cos^2 \theta)},
\]
\[
\sin \phi = -\frac{27}{(5 - 4 \cos \theta)(157 + 44 \cos \theta - 32 \cos^2 \theta)}.
\]
Plots of $\cos \phi$ and $\sin \phi$ as a function of $\theta$ are shown in Figure 6. For $s = e^{i\theta}$, Eq. (22) is equivalent to $\cos((k-1)\theta) = \cos \phi$ and $\sin((k-1)\theta) = \sin \phi$. Given that $1/s = e^{-i\theta}$, we can focus on solutions such that $\theta \in [0, \pi]$. The oscillations of $\cos((k-1)\theta)$ in this interval imply that the equation $\cos((k-1)\theta) = \cos \phi$ admits at least one solution in each interval $\left(\frac{m-1}{k}\pi, \frac{m+1}{k}\pi\right)$, for $1 \leq m \leq k-1$. For each solution, $\sin((k-1)\theta) = \pm \sin \phi$, and the plot of $\sin \phi$ in Figure 6 shows that $\sin((k-1)\theta) = \sin \phi$ if and only if $\sin((k-1)\theta) \leq 0$, that is, if $m$ is even. We finally note that, when $k$ is odd, one solution is $\theta = \pi$, giving $s = -1$, which we want to exclude.

This discussion shows that (22) has at least $\left\lceil \frac{k-1}{k} \right\rceil$ solutions $s \neq -1$ with $\text{Im}(s) > 0$ on the unit circle. They give rise to as many roots of $G_k$ in the interval $[-1, -1/9]$. With the two real roots of $G_k$ found previously outside this interval, this gives a total of $\left\lceil \frac{k-1}{k} \right\rceil$ roots, which coincides with the degree of $G_k$. Hence $G_k$ is real rooted, with one positive root $\rho_k$, and the others smaller than $-1/9$.

In remains to obtain an expansion of $\rho_k$ as $k$ grows. We first work out an expansion of the solution of (22) found around $1/2$ (by bootstrapping in (22)):
\[
s = \frac{1}{2} - \frac{3}{2k^2} - \frac{3}{4k^2} + \frac{36k + 27}{4,8k^4} - \frac{27 \cdot 32k^2 - 9 \cdot 16k - 717}{16k^2} + O\left(\frac{k^3}{32^{k}}\right).
\]
This translates into the expansion of $\rho_k$ using (23). The singular behaviour of $M_k$ is then derived from (20).

\[\text{Step 4. Conclusion}\]

By Lemma 7 and Cauchy’s formula,
\[
[x^r] \left( M_k(x) - \frac{\alpha_k}{1 - x/\rho_k} \right) = \frac{1}{2i\pi} \oint_C \left( M_k(x) - \frac{\alpha_k}{1 - x/\rho_k} \right) \frac{dx}{x^{r+1}}, \quad (24)
\]
where $C$ is the circle of radius $1/9$ centered at the origin. We will prove that there exists a constant $C$ such that for all $k$ and $x \in C$,

$$|M_k(x) - \frac{\alpha_k}{1 - x/\rho_k}| \leq Ck,$$

so that (24) implies

$$[x^t]M_k(x) = \mathbb{E}(X_{x,t}^2) = \alpha_k\rho_k^{-t} + O(9^tk),$$

as stated in Proposition 5.

It follows from (14) and (16) that $\frac{\alpha_k}{1 - x/\rho_k}$ is bounded uniformly in $k$ and $x \in C$, so that we only need to prove that $M_k(x) = O(k)$, uniformly in $x \in C$. By Lemma 6, for any $x \in C \setminus \{-1/9\}$, $|S(x)| < 1$. Moreover, $S(x) \to 1$ as $x \to -1/9$. Recall that $M_k(x) = 2xN_k/G_k$. By (17), $N_k = 2^{k/2}O(k)$, so that it suffices to prove that $G_k/2^{k/2}$ is bounded away from 0, uniformly in $k$ and $x \in C$. Since $1 + S$ and $S^k$ are uniformly bounded, and $P(1/S)$ is bounded away from 0, this is equivalent to

$$\inf_{k, x \in C} \left| S^k + \frac{P(S)}{P(1/S)} \right| > 0.$$

By the proof of Lemma 7, $S^k + \frac{P(S)}{P(1/S)}$ does not vanish on $C$. Hence it suffices to prove that

$$\liminf_{k} \inf_{x \in C} \left| S^k + \frac{P(S)}{P(1/S)} \right| > 0.$$

Let us write $x = -e^{\pm i\theta}/9$, with $\theta \in [0, \pi]$. Then, as $\theta \to 0$,

$$S(x) = 1 - \frac{1}{2}(1 \mp i)\sqrt{\theta} + O(\theta), \quad |S(x)| = 1 - \frac{1}{2}\sqrt{\theta} + O(\theta), \quad \frac{P(S)}{P(1/S)} = 1 - \frac{20}{13}(1 \mp i)\sqrt{\theta} + O(\theta).$$

We split then interval $[0, \pi]$, to which $\theta$ belongs, in three parts.

- When $\sqrt{\theta} \leq \pi/(2k)$, there holds, uniformly in $\theta$,

$$S(x)^k = \exp\left(-k(1 \mp i)\sqrt{\theta}/2\right) + O(1/k).$$

In particular,

$$\Re(S^k) = \exp(-k\sqrt{\theta}/2)\cos(k\sqrt{\theta}/2) + O(1/k) \geq \exp(-\pi/4)/\sqrt{2} + O(1/k).$$

Moreover,

$$\Re\left(\frac{P(S)}{P(1/S)}\right) = 1 + O(\sqrt{\theta}) = 1 + O(1/k),$$

uniformly in $\theta$. Hence

$$\Re\left(S^k + \frac{P(S)}{P(1/S)}\right) = 1 + \exp(-\pi/4)/\sqrt{2} + O(1/k),$$

and

$$\liminf_{k} \inf_{\sqrt{\theta} \leq \pi/(2k)} \left| \frac{S^k + P(S)}{P(1/S)} \right| > 0.$$

- Let $\varepsilon > 0$ be such that, for $\sqrt{\theta} < \varepsilon$,

$$|S(x)| \leq 1 - \sqrt{\theta}/4 \quad \text{and} \quad \left| \frac{P(S)}{P(1/S)} \right| > 0.9.$$
Such an $\epsilon$ exists in view of (26) and (27). For $\pi/(2k) \leq \sqrt{\theta} \leq \epsilon$, $|S|^k \leq \left(1 - \sqrt{\theta}/4\right)^k \leq \left(1 - \pi/(8k)\right)^k = \exp(-\pi/8) + O(1/k)$, so that $\left|S^k + \frac{P(S)}{P(1/S)}\right| \geq 0.9 - \exp(-\pi/8) + O(1/k) \geq 0.2 + O(1/k)$.

Hence
\[
\lim \inf_k \inf_{\pi/(2k) \leq \sqrt{\theta} \leq \epsilon} \left|S^k + \frac{P(S)}{P(1/S)}\right| > 0.
\]

Finally, when $\sqrt{\theta} \geq \epsilon$, then $|S| < 1$ is bounded away from 1, uniformly in $\theta$. Thus, if
\[
\lim \inf_k \inf_{\sqrt{\theta} \geq \epsilon} \left|S^k + \frac{P(S)}{P(1/S)}\right| = 0,
\]
there would exist an $x \in C$ such that $P(S(x)) = 0$. But this only happens when $x = 0$ or $x = (-1 \pm \sqrt{\theta})/4$, and none of these values lies on the circle $C$.

This concludes the proof of Proposition 5.

\section*{4. Back to Knuth's algorithm}

Let us go back to Knuth's original algorithm, described at the beginning of the paper. Recall that $E(X_k)$ is the number of SAWs crossing a square of side $k$, and that $E(X_k^2)$ is the sum of the reciprocals of the probabilities of these walks.

\begin{proposition}
Denote $c(k) = E(X_k)$ and $d(k) = E(X_k^2)$. There exist two positive constants $\lambda$ and $\beta$ such that
\[
E(X_k^{1/k^2}) \to \lambda \quad \text{and} \quad E(X_k^{2/(k+1)}) \to \beta.
\]

Of course, $\beta \geq \lambda^2$. Moreover,
\[
\lambda = \sup_k c(k)^{1/(k+1)^2}
\]
and
\[
\beta = \sup_k \left(\sqrt{d(k)}\right)^{1/(k+1)^2}.
\]
\end{proposition}

As discussed at the end of the introduction, there is a hope to combine (29) and known upper bounds on $\lambda$ to prove that $\beta > \lambda^2$, in which case one could conclude that the relative variance of $X_k$ grows as $k^{2\beta}$, with $\kappa = \beta/\lambda^2 > 1$.

\begin{proof}
As can be expected, these results follow from a super-multiplicativity argument. The existence of $\lambda$ was established for the first time in [1], and (28) (which allows to produce lower bounds on $\lambda$) appears in [7]. We repeat the argument, because it applies almost verbatim to the numbers $d(k)$.

Define $\lambda := \lim \sup_k c(k)^{1/k^2}$. Then $\lambda$ is finite, because there are only a quadratic number of edges in the $k \times k$ square, and a walk is determined by the set of its edges. Let $\epsilon > 0$. We will prove that
\[
\lim \inf c(K)^{1/K^2} \geq \lambda - \epsilon,
\]
which implies that $\lambda$ is actually the limit of $c(k)^{1/k^2}$.

Let $k > 0$ be such that $c(k)^{1/(k+1)^2} > \lambda - \epsilon$. Let $K \geq k$, and let $n$ be maximal so that
\[
(k + 1)(2n + 1) - 1 \leq K.
\]
This implies in particular that $K < (k + 1)(2n + 3)$. In the $K \times K$ square, put $(2n + 1)^2$ smaller squares of side $k$, as shown in Figure 7. In each smaller square, choose a SAW that crosses it, and build from this collection of short walks a long walk crossing the larger square, as shown in the figure. This construction implies
\[
c(K) \geq c(k)^{(2n+1)^2}.
\]
Thus
\[ c(K)^{1/K^2} \geq c(K)^{1/((k+1)^2(2n+3)^2)} \geq \left( c(k)^{1/(k+1)^2} \right)^{(2n+1)^2/(2n+3)^2} \geq (\lambda - \varepsilon)^{(2n+1)^2/(2n+3)^2}. \]

Taking the lim inf on \( K \) boils down to taking the lim inf on \( n \) and gives (30). The bound (28) also follows from the above inequalities.

\[ \text{Figure 7. Super-multiplicativity for SAWs crossing a square} \]

Let us now consider the numbers \( d(k) \). Again, \( \beta := \limsup d(k)^{1/k^2} \) is finite, because
\[ d(k) = \sum_{w \in W_k} \frac{1}{p(w)} \leq \sum_{w \in W_k} 3^{|w|} \leq 3^{O(k^2)} c(k), \]
where \( |w| \) denotes the length of \( w \). Now return to Figure 7. Denote by \( w_1, w_2, \ldots \) the short walks, and by \( w \) the long one. It is clear for the sampling algorithm that
\[ \frac{1}{p(w)} \geq \prod_{i=1}^{(2n+1)^2} \frac{1}{p(w_i)}. \]
It follows that
\[ d(K) \geq d(k)^{(2n+1)^2}, \]
from which one can prove, as above, that \( \beta = \lim d(k)^{1/k^2} \). The above bound on \( d(K) \) can actually be improved: in every row of \( (2n+1)^2 \) small squares, except maybe the top one, \( n \) of the horizontal steps added between the small squares have probability 1/2 or 1/3. Hence
\[ d(K) \geq 2^{2n^2} d(k)^{(2n+1)^2}, \]
and the lower bound (29) now follows.

5. Final comments

5.1. Unconfined walks

Knuth designed his algorithm to sample SAWs crossing a square of side \( k \), but other authors have used similar ideas to sample unconfined SAWs of fixed length \( n \). For example, the classical Rosenbluth algorithm [21] generates general SAWs step by step, by taking at each time, uniformly at random, one of the steps that preserves self-avoidance. If at some point no such step is available, and the walk has not reached length \( n \), the algorithm restarts from scratch. This rejection step is avoided if one only samples untrapped self-avoiding walks, that is, walks that can
be extended into a SAW of infinite length. (We describe in Section 5.3 a simple procedure that detects if a new step traps the walk, which is also useful for implementing Knuth’s algorithm.) A recent numerical study, using a refinement of the above algorithm, suggests that the asymptotic properties of untrapped walks are similar to those of general SAWs, in terms of number and end-to-end distance.

For these algorithms, the quality of the cardinality estimator is still related to the variance of the random variable $X_n$ equal to the reciprocal of the probability of the generated walk. As already mentioned, the variance of the Rosenbluth estimator is predicted to be exponential in $n$ [6]. We do not know of any similar study for untrapped walks. It is easy to determine the variance of $X_n$ for the Rosenbluth algorithm restricted to directed or partially directed walks. Our results are summarized in the following table. In particular, for partially directed walks the relative variance is found to be exponential in $n$.

<table>
<thead>
<tr>
<th></th>
<th>North and East</th>
<th>North, East and South</th>
<th>all four steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>confined</td>
<td>number</td>
<td>$4^k/\sqrt{k}$</td>
<td>$(k+1)^k$</td>
</tr>
<tr>
<td>to $k \times k$ rel. var.</td>
<td>$\sqrt{k}$ [5]</td>
<td>$2^{k(k+1)}/(k+1)^k$ (Prop. 5)</td>
<td>$\lambda^k$</td>
</tr>
<tr>
<td>av. length</td>
<td>$k$</td>
<td>$k^2$</td>
<td>$k^2$ [16]</td>
</tr>
<tr>
<td>unconfined,</td>
<td>number</td>
<td>$2^n$</td>
<td>$(1+\sqrt{2})^n$</td>
</tr>
<tr>
<td>$n$ steps</td>
<td>rel. var.</td>
<td>0</td>
<td>$(6/(1+\sqrt{2}))^n$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\alpha^n$ (pred. [6])</td>
</tr>
</tbody>
</table>

### 5.2. Kinetic Distributions

It is also interesting to study the asymptotic properties of SAWs chosen according to the non-uniform (but very natural) “kinetic” distribution that results from importance sampling. These properties may be different from those observed in the uniform case. For instance, one can expect the average end-to-end distance of kinetic unconfined SAWs to be smaller than $n^{3/4}$, because “compact” walks in which few steps are eligible at each time have a higher probability than more spread-out walks. In fact, the kinetic end-to-end distance is conjectured [18] to grow like $n^{2/3}$. (For unconfined partially directed walks, however, the end-to-end distance is easily shown to be linear, both for the uniform and the kinetic model.) Figure 8 shows a random (untrapped) SAW that we generated by importance sampling and a (quasi-)uniform SAW generated using a pivot algorithm [13].

![Figure 8. A random untrapped SAW of length 5000 obtained via importance sampling (left), and a quasi-uniform SAW of length 20000 (right).](image)

---

4This notion of untrapped walks differs from the one in [19], where a walk is said to be untrapped as soon as it can be extended by one step.
5.3. When does a walk get trapped?

One important feature of Knuth’s algorithm, and of its adaptation to untrapped SAWs discussed in Section 5.1, is that one never appends a step that would trap the walk. Since Knuth does not explain in his paper how he detects trapping, let us describe the method we used. One obvious case of trapping in Knuth’s algorithm is when the walk reaches the boundary of the square, and moves towards the origin. In all other trapping situations, the walk would have been trapped as well in the unconfined setting, so we focus on the trapping of unconfined walks.

Let $w$ be an untrapped SAW of length $n$, ending at vertex $v_n = (i, j)$, and, say, with a $W$ step. There are, up to obvious symmetries, exactly three situations when adding a new step to $w$ creates a trapped walk:

- the vertex $v = (i - 1, j)$ belongs to $w$, one appends a $N$ step to $w$ and the portion of $w$ going from $v$ to $v_n$ has winding number $-2\pi$,
- the vertex $v = (i - 1, j + 1)$ belongs to $w$, one appends a $N$ step to $w$ and the portion of $w$ going from $v$ to $v_n$ has winding number $-2\pi$,
- the vertex $v = (i - 1, j + 1)$ belongs to $w$, one appends a $W$ or $S$ step to $w$ and the portion of $w$ going from $v$ to $v_n$ has winding number $2\pi$.

These three cases are depicted in Figure 9. When computing the winding number, we add a half-edge pointing from the East to $v$ (Figure 10). The winding number is then the difference between the number of left turns and the number of right turns, multiplied by $\pi/2$.

![Figure 9. How a walk gets trapped.](image1)

![Figure 10. The winding number between $v$ and $v_n$ is $-2\pi$.](image2)

Acknowledgements. I am very grateful to Persi Diaconis, who first told me about this question, and then sent helpful suggestions on this manuscript and pointed to interesting related papers. I also thank the referee of a first version of this paper for his very thorough report and helpful references.

References

ON THE IMPORTANCE SAMPLING OF SELF-AVOIDING WALKS


MBM: CNRS, LABRI, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence, France
E-mail address: mireille.bousquet@labri.fr