

A Combinatorial Approach to a Model of Constrained Random Walkers

T. ESPINASSE, N. GUILLOTIN-PLANTARD and P. NADEAU

Institut Camille Jordan, CNRS UMR 5208, Université Lyon 1, 43 Boulevard du 11 novembre 1918,
69622 Villeurbanne, France

(e-mail: Thibault.Espinasse@math.univ-lyon1.fr,
nadine.guillotin@univ-lyon1.fr,
philippe.nadeau@math.univ-lyon1.fr)

Received 24 March 2014; revised 11 December 2014; first published online 16 March 2015

In [1], the authors consider a random walk $(Z_{n,1}, \dots, Z_{n,K+1}) \in \mathbb{Z}^{K+1}$ with the constraint that each coordinate of the walk is at distance one from the following coordinate. A functional central limit theorem for the first coordinate is proved and the limit variance is explicit. In this paper, we study an extended version of this model by conditioning the extremal coordinates to be at some fixed distance at every time. We prove a functional central limit theorem for this random walk. Using combinatorial tools, we give a precise formula of the variance and compare it with that obtained in [1].

2010 *Mathematics subject classification*: Primary 60J10
Secondary 60F05, 05A19

1. Introduction and results

Central limit theorems for additive functionals of Markov chains have attracted continuing interest for over half a century. One approach, due to Gordin [3], rests on martingale approximation. Roughly speaking, in good cases the asymptotic normality for functionals of Markov chains is derived from the central limit theorem for martingales. Various conditions for the central limit theorem to hold are now known and several expressions for the asymptotic variance can be found in the literature (see for instance [4]), but it is usually quite difficult to compute it theoretically. In this note, we are faced with this problem in the study of a generalized version of the so-called ‘prisoners model’ introduced in [1]. We are able to give an explicit formula for the asymptotic variance of the Markov chain we are interested in. However, our approach is not sufficiently robust to give a full description of the asymptotic variance for more general models (see remark (3) in Section 5).

In what follows we will use the notation $[n] := \{1, \dots, n\}$. Let K be a positive integer. For any $h \in \{0, \dots, K\}$, we define

$$C_{K,h} = \{z \in \mathbb{Z}^{K+1}; \forall i \in [K], |z_{i+1} - z_i| = 1 \text{ and } z_{K+1} - z_1 = h\}.$$

The set $C_{K,h}$ is empty unless $K - h$ is even, so let us set $K - h = 2g$. The set $C_{K,0}$ corresponds to Bernoulli bridges with length K .

We can define a neighbourhood structure on $C_{K,h}$ via the following relation:

$$\forall z, z' \in C_{K,h}, z \sim z' \Leftrightarrow \forall i \in [K + 1], |z_i - z'_i| = 1.$$

The set of neighbours of $z \in C_{K,h}$ will be denoted by $\Gamma(z)$ and its cardinality by $\text{deg}_{C_{K,h}}(z)$.

Let $(Z_n)_{n \geq 0}$ denote the Markov chain defined on $C_{K,h}$ corresponding to $K + 1$ simple random walks on \mathbb{Z} under the shape constraint. In other words, $(Z_n)_{n \geq 0}$ is the Markov chain with state space $C_{K,h}$ and transition probabilities given by

$$\mathbb{P}[Z_{n+1} = z' | Z_n = z] = \begin{cases} \frac{1}{\text{deg}_{C_{K,h}}(z)} & \text{if } z' \sim z, \\ 0 & \text{otherwise.} \end{cases}$$

We will denote by $Z_{n,i}$ the i th coordinate of Z_n and by $[x]$ the integer part of a real number x . We assume that $Z_{0,1} = 0$ almost surely.

In this paper we are interested in the distributional limit of the first coordinate $(Z_{n,1})$ as n tends to infinity. Actually, the convergence to Brownian motion is not surprising, and we are mainly interested here in the exact value of the variance.

Theorem 1.1. *The sequence of random processes*

$$\left(\frac{Z_{[nt],1}}{\sqrt{n}} \right)_{t \geq 0, n \geq 1}$$

weakly converges to Brownian motion with variance

$$\sigma_{K,h}^2 = \frac{1}{K} \frac{A_{K,h}}{B_{K,h}},$$

where

$$A_{K,h} = \sum_{k=0}^{\lfloor K/2 \rfloor} \binom{K}{2k+1} \binom{K-2k}{g-k}$$

and

$$B_{K,h} = \sum_{k=0}^{\lfloor K/2 \rfloor} \binom{K}{2k} \binom{K-2k}{g-k}.$$

The variance $\sigma_{K,h}^2$ has the following properties.

(i) For any $h \in \{0, \dots, K\}$, as K tends to infinity,

$$\sigma_{K,h}^2 \sim \frac{2}{K}.$$

(ii) For any K ,

$$\sigma_{K,0}^2 < 2/K.$$

(iii) For K large enough,

$$\sigma_{K,0}^2 > 2/(K + 2).$$

With our notations, the variance obtained in [1] is equal to $2/(K + 2)$, so item (iii) means that for large K our variance is larger than that obtained in the unconstrained case. We conjectured that this should hold for any K , and it has been shown independently by Kilian Raschel, Oriane Blondel, and François Chapon. We warmly thank them for their help.

The paper is organized as follows. In Section 2, various enumerative results about walks are proved. They will be used in Section 3. Theorem 1.1 is proved in Section 3. Our approach is based on a decomposition of $Z_{n,1}$ as a sum of a martingale and a bounded function. We will compute the two parts explicitly, and provide a geometrical interpretation for the decomposition. The asymptotic properties of the variance will be proved using local limit theorems for random walks on \mathbb{Z} . In Section 4 we explain how to adapt our method to derive the main result of [1]. In Section 5 some variations of the model are proposed and open problems are discussed.

2. Combinatorics

We prove various enumerative results that will be used in Section 3.

If $z \in \mathcal{C}_{K,h}$, its shape is defined as $\mathcal{F}(z) = (F_1, \dots, F_K)$ by $F_i = z_{i+1} - z_i$. We refer naturally to the elements of $\mathcal{F}(z)$ as the *steps* of z , which can be either up or down. The set of possible shapes is given by

$$Sh_{K,h} = \left\{ (F_i)_i \in \{\pm 1\}^K ; \sum_{i=1}^K F_i = h \right\}.$$

Clearly $z \in \mathcal{C}_{K,h}$ if and only if $(z_1, \mathcal{F}(z)) \in \mathbb{Z} \times Sh_{K,h}$. Let us denote, for every $z \in \mathcal{C}_{K,h}$,

$$\Gamma^\pm(z) = \{z' \sim z ; z'_1 - z_1 = z'_{K+1} - z_{K+1} = \pm 1\}.$$

Now let $z, z' \in \mathcal{C}_{K,h}$ be such that $z' \in \Gamma^+(z)$, and let $F = \mathcal{F}(z), F' = \mathcal{F}(z')$ be the shapes of z, z' respectively. For any $i \in [K]$, define $M_i = \pm 1$ if $F_i = F'_i = \pm 1$ and $M_i = 0$ otherwise. So $M = (M_1, \dots, M_K)$ has values in $\{1, 0, -1\}$, and can thus be represented as a Grand Motzkin path; we define $\Phi^+(z, z') = M$. Note that $\sum_i M_i = h$, so this path M goes from $(0, 0)$ to (K, h) ; we let $\mathcal{M}_{K,h}$ be the set of such sequences, which has cardinality

$$|\mathcal{M}_{K,h}| = \sum_{k=0}^{\lfloor K/2 \rfloor} \binom{K}{2k} \binom{K-2k}{g-k}. \tag{2.1}$$

Now if we fix the starting point of z , then from any element of $\mathcal{M}_{K,h}$ we can clearly reconstruct a pair z, z' , so we have the following result.

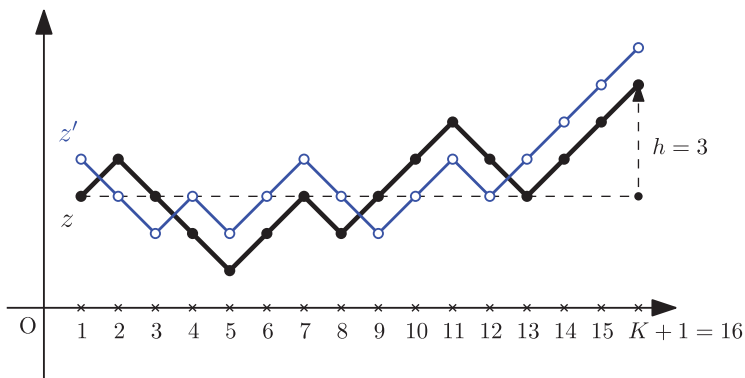


Figure 1. Two paths $z \sim z'$ with $z' \in \Gamma^+(z)$.

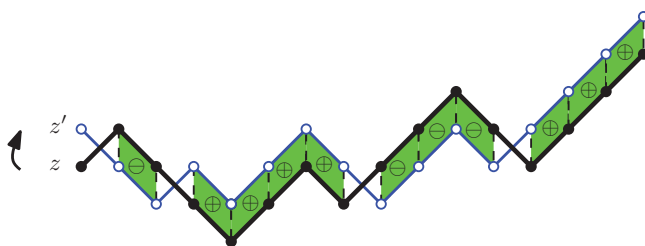


Figure 2. The area $A(z') - A(z)$ is equal to $-1 + 4 - 3 + 3 = 3$.

Proposition 2.1. For any z_1 , Φ^+ is a bijection between

$$\{(z, z') \in \mathcal{C}_{K,h}^2; z' \in \Gamma^+(z)\}$$

and $\mathcal{M}_{K,h}$.

Let $A(z)$ denote the algebraic area between the path z and the X -axis. Note that

$$A(z) = \frac{1}{2}z_1 + z_2 + \dots + z_K + \frac{1}{2}z_{K+1}.$$

We now prove that for any fixed path $z \in \mathcal{C}_{K,h}$ the sum over $z' \in \Gamma(z) := \Gamma^+(z) \cup \Gamma^-(z)$ of the algebraic areas boils down to zero. Fix a path $z \in \mathcal{C}_{K,h}$. Let (z', p) be a *marked path* (with respect to z), which we define as a path $z' \in \Gamma(z)$ together with a step $p \in z'$ which does not cross z . Define $\text{sign}(z; z', p) = 1$ (resp. -1) if the step p occurs above (resp. below) z . With these notations it is clear that $A(z') - A(z) = \sum_p \text{sign}(z; z', p)$, where the sum is over all marking steps in z' ; see Figure 2.

We now define an involution \mathcal{I} of these marked paths. Note first that z' crosses z an even number of times, and let $2k$ be this quantity. The case $k = 0$ occurs only when $z' \in \{z^+, z^-\}$, which are defined as the two paths obtained from z by shifting it one unit up or down. In this case simply define $\mathcal{I}(z; z^\pm, p) = (z; z^\mp, p)$. Suppose now that $k \neq 0$, and let L (resp. R) denote the nearest crossing left (resp. right) of p . There are two special cases: if p occurs before the first crossing of z' , then L is defined as the last crossing of z' ,

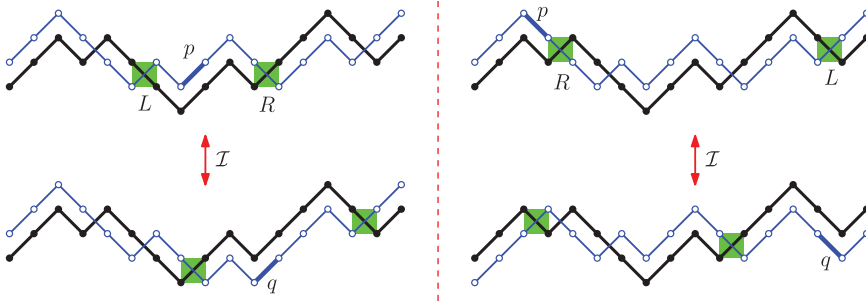


Figure 3. Two cases of the application of the involution \mathcal{I} .

while if p occurs after the last crossing of z' , then R is defined as the first crossing of z' . Examples of the definition of L, R are illustrated in Figure 3.

Now exactly one step among $\{L, R\}$ is a step of z' of a different up/down type to p . Let q be this crossing step, and exchange p with q in z' , keeping all other steps unchanged. This defines a new path $z'' \in \Gamma(z)$, in which p becomes crossing while q is non-crossing. So (z'', q) is a marked path, and we define $\mathcal{I}(z; z', p) := (z; z'', q)$. We have the following lemma, whose proof is immediate.

Lemma 2.2. For any z , \mathcal{I} is an involution on marked paths, which is sign reversing.

Proposition 2.3. For any $z \in \mathcal{C}_{K,h}$,

$$\sum_{z' \in \Gamma(z)} (A(z') - A(z)) = 0.$$

Proof. The sum to compute is equal to the sum of $\text{sign}(z; z', p)$ for all marked paths (z', p) . By the previous lemma, the involution \mathcal{I} pairs such marked paths two by two with opposite signs, so the sum boils down to zero. \square

Proposition 2.4. For any K, h ,

$$\sum_{\substack{z, z' \in \mathcal{C}_{K,h} \\ z' \in \Gamma^+(z)}} [A(z') - A(z)] = \sum_{k=0}^{\lfloor K/2 \rfloor} \binom{K}{2k+1} \binom{K-2k}{g-k}.$$

Proof. The left-hand side can be rewritten as the sum of $\text{sign}(z; z', p)$ over all z, z', p such that $z' \in \Gamma^+(z)$ and (z', p) is marked. Define $(z, z'', q) := \mathcal{I}(z, z', p)$. If z'' also belongs to $\Gamma^+(z)$, then the corresponding terms in the sum will cancel each other. The surviving terms correspond to $z'' \in \Gamma^-(z)$, which by inspection occurs precisely in the following cases:

- (i) p occurs before the first crossing, and p is a descent;
- (ii) p occurs after the last crossing, and p is an ascent;

(iii) $z' = z^+$ and p is any step.

In all three cases, $\text{sign}(z, z', p)$ is always equal to 1. Thus we need to enumerate the triplets $(z; z', p)$ verifying (i), (ii) or (iii).

There are clearly $K \binom{K}{g}$ triplets satisfying (iii). We compute the number of triplets (z, z', p) verifying case (i) or (ii), and such that there are $2k > 0$ crossings between z and z' . The position of p together with the position of these crossings form a subset $\{j_1 < j_2 < \dots < j_{2k+1}\}$ of $[K]$; here the position of p is j_1 in case (i) and j_{2k+1} in case (ii). There are $\binom{K}{2k+1}$ such subsets, and, conversely, the knowledge of such a subset determines the positions of p and the crossings of z and z' in each case. The remaining $K - 2k - 1$ steps are the same in z and z' , so we only need to count possible steps for z . We want to ensure that z belongs to $\mathcal{C}_{K,h}$. Notice that among the crossings there are k up-steps and k down-steps. Now in case (i), p is a down-step, so there must be exactly $g - k - 1$ down-steps among the remaining $K - 2k - 1$ steps; in case (ii), p is an up-step, so there must be exactly $g - k$ down-steps among the remaining $K - 2k - 1$ steps. In total, this represents

$$\binom{K - 2k - 1}{g - k - 1} + \binom{K - 2k - 1}{g - k} = \binom{K - 2k}{g - k}$$

possibilities.

Adding everything up, we obtain the desired expression (note that the case $k = 0$ corresponds to (iii)). □

3. Proof of Theorem 1.1

3.1. Martingale

Let $(\mathcal{F}_n)_{n \geq 0}$ denote the natural filtration of the Markov chain Z_n .

Lemma 3.1. *The sequence $(A(Z_n))_{n \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$.*

Proof. Since $(Z_n)_n$ is a Markov chain, it is enough to prove that for any $z \in \mathcal{C}_{K,h}$,

$$\mathbb{E}[A(Z_{n+1}) - A(Z_n) | Z_n = z] = 0,$$

or equivalently (since the Markov chain moves uniformly on its neighbours) for any $z \in \mathcal{C}_{K,h}$,

$$\sum_{z' \sim z} [A(z') - A(z)] = 0.$$

We recognize here the content of Proposition 2.3. □

3.2. The coupling

Let $f_K(z)$ denote the area between a path $z = (z_1, z_2, \dots, z_K, z_{K+1}) \in \mathcal{C}_{K,h}$ and the segment $(z_1, z_1 + 1, \dots, z_1 + K)$. Then, for any $z \in \mathcal{C}_{K,h}$ (see Figure 4),

$$f_K(z) + A(z) = Kz_1 + \frac{K^2}{2}. \tag{3.1}$$

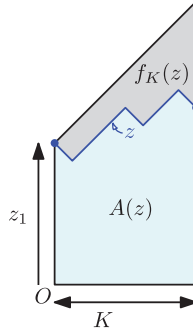


Figure 4. Relation between the first coordinate z_1 and the area under the path.

Since f_K is bounded by K^2 , the position of the first walker and the area under the path can be asymptotically related as

$$\frac{Z_{[nt],1}}{\sqrt{n}} = \frac{1}{K} \frac{A(Z_{[nt]})}{\sqrt{n}} + o(1). \tag{3.2}$$

Theorem 1.1 will then be deduced from the functional central limit theorem for the rescaled martingale $(A(Z_n)/\sqrt{n})_{n \geq 1}$.

If $z \in \mathcal{C}_{K,h}$, its shape is defined as $\mathcal{F}(z) = (F_1, \dots, F_{K+1})$ by $F_i = z_i - z_1$. The set of possible shapes is given by

$$\overline{\mathcal{S}h}_{K,h} = \{z \in \mathcal{C}_{K,h}; z_1 = 0\}.$$

This description of the possible shapes of $\mathcal{C}_{K,h}$ differs from the one used in Section 2 but will be more convenient in the next computations. Note that a natural neighbourhood structure is given on $\overline{\mathcal{S}h}_{K,h}$ by the relation

$$F \sim_S F' \Leftrightarrow \forall i \in [K + 1], F_i - F'_i \in \{0, 2\} \quad \text{or} \quad \forall i \in [K + 1], F_i - F'_i \in \{0, -2\}.$$

If F, F' are neighbours in $\overline{\mathcal{S}h}_{K,h}$, then one of following statements is true.

- If $\forall i \in [K + 1], F_i - F'_i \in \{0, 2\}$ then $F - (1, \dots, 1)$ and F' are neighbours in $\mathcal{C}_{K,h}$.
- If $\forall i \in [K + 1], F_i - F'_i \in \{0, -2\}$ then $F + (1, \dots, 1)$ and F' are neighbours in $\mathcal{C}_{K,h}$.

Note that the special case $F = F'$ gives two neighbours in $\mathcal{C}_{K,h}$, while the other cases provide a one-to-one transformation.

Conversely, if z, z' are neighbours in $\mathcal{C}_{K,h}$, then either $\mathcal{F}(z) = \mathcal{F}(z')$ and $z = z' \pm (1, \dots, 1)$, or $\mathcal{F}(z)$ and $\mathcal{F}(z')$ are distinct neighbours in $\overline{\mathcal{S}h}_{K,h}$.

We denote by $\text{deg}_S(F)$ the number of neighbours of $F \in \overline{\mathcal{S}h}_{K,h}$, that is,

$$\text{deg}_S(F) := |\{G \in \overline{\mathcal{S}h}_{K,h}; G \sim_S F\}|.$$

According to the previous remark, we have

$$\text{deg}_{\mathcal{C}_{K,h}}(z) = \text{deg}_S(\mathcal{F}(z)) + 1.$$

Consider the Markov chain $(F_n)_n$ with values in $\overline{Sh}_{K,h}$, with transition probabilities given by

$$p(F, G) = \begin{cases} \frac{1}{\deg_S(F) + 1} & \text{if } G \sim_S F \text{ and } G \neq F, \\ \frac{2}{\deg_S(F) + 1} & \text{if } G = F. \end{cases}$$

This irreducible and ergodic Markov chain is reversible with stationary measure

$$\pi_S(F) := \frac{\deg_S(F) + 1}{\sum_{F \in \overline{Sh}_{K,h}} (\deg_S(F) + 1)}.$$

The Markov chains $(\mathcal{F}(Z_n))_{n \in \mathbb{N}}$ and $(F_n)_{n \in \mathbb{N}}$ are then identically distributed. Indeed,

$$\begin{aligned} \mathbb{P}[\mathcal{F}(Z_{n+1}) = G | Z_n = z] &= \frac{1}{\deg_{\mathcal{C}_{K,h}}(z)} \sum_{z' \sim z} 1_{\mathcal{F}(z')=G} \\ &= \begin{cases} \frac{1}{\deg_S(\mathcal{F}(z)) + 1} & \text{if } G \sim_S \mathcal{F}(z), G \neq \mathcal{F}(z) \\ \frac{2}{\deg_S(\mathcal{F}(z)) + 1} & \text{if } G = \mathcal{F}(z) \end{cases} \\ &= \mathbb{P}[F_{n+1} = G | F_n = \mathcal{F}(z)]. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbb{P}[\mathcal{F}(Z_{n+1}) = G | \mathcal{F}(Z_n) = F] \\ &= \sum_{z \in \mathcal{C}_{K,h}} \mathbb{P}[\mathcal{F}(Z_{n+1}) = G | \mathcal{F}(Z_n) = F; Z_n = z] \mathbb{P}[Z_n = z | \mathcal{F}(Z_n) = F] \\ &= \mathbb{P}[F_{n+1} = G | F_n = F] \left(\sum_{z \in \mathcal{C}_{K,h}} \mathbb{P}[Z_n = z | \mathcal{F}(Z_n) = F] \right) \\ &= \mathbb{P}[F_{n+1} = G | F_n = F]. \end{aligned}$$

Let us remark that for every $z, z' \in \mathcal{C}_{K,h}$ such that $z' \sim z$ with $z' \neq z$, the difference in area $|A(z') - A(z)|$ can be computed with respect to $\mathcal{F}(z')$ and $\mathcal{F}(z)$ only. Indeed, we have the following relations:

$$|A(z') - A(z)| = \begin{cases} |A(\mathcal{F}(z')) - A(\mathcal{F}(z) + (1, \dots, 1))| & \text{if } \exists i \text{ s.t. } \mathcal{F}(z')_i - \mathcal{F}(z)_i = 2, \\ |A(\mathcal{F}(z')) - A(\mathcal{F}(z) - (1, \dots, 1))| & \text{if } \exists i \text{ s.t. } \mathcal{F}(z')_i - \mathcal{F}(z)_i = -2, \\ K & \text{if } \mathcal{F}(z') = \mathcal{F}(z). \end{cases}$$

The quantity $A(z') - A(z)$ will then be denoted by $\Delta A(\mathcal{F}(z), \mathcal{F}(z'))$. From the central limit theorem for martingales (see for instance [2, Theorem 7.4, p. 374]), we have to compute

the a.s. limit as n tends to infinity of

$$\begin{aligned} \Sigma_n &:= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[(A(Z_k) - A(Z_{k-1}))^2 | Z_{k-1}] \\ &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\Delta A(\mathcal{F}(Z_{k-1}), \mathcal{F}(Z_k))^2 | \mathcal{F}(Z_{k-1}); Z_{k-1,1}] \\ &= \sum_{F \in \overline{S}h_{K,h}} \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{\mathcal{F}(Z_{k-1})=F\}} \mathbb{E}[\Delta A(\mathcal{F}(Z_{k-1}), \mathcal{F}(Z_k))^2 | \mathcal{F}(Z_{k-1}) = F] \\ &= \sum_{F \in \overline{S}h_{K,h}} \left(\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{\mathcal{F}(Z_{k-1})=F\}} \right) \mathbb{E}[\Delta A(\mathcal{F}(Z_0), \mathcal{F}(Z_1))^2 | \mathcal{F}(Z_0) = F] \\ &\stackrel{\text{law}}{=} \sum_{F \in \overline{S}h_{K,h}} \left(\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{F_{k-1}=F\}} \right) \mathbb{E}[\Delta A(F_0, F_1)^2 | F_0 = F]. \end{aligned}$$

Since the Markov chain $(F_n)_n$ is ergodic with invariant measure π_S , the sequence of random variables $(\Sigma_n)_n$ converges almost surely to the constant

$$\sum_{F \in \overline{S}h_{K,h}} \mathbb{E}[(\Delta A(F_0, F_1))^2 | F_0 = F] \pi_S(F),$$

which can be rewritten as $\mathbb{E}_{\pi_S} [(A(Z_1) - A(Z_0))^2]$ by noting that $\mathcal{F}(Z_0) = Z_0$. Then, using (3.2), we obtain that

$$\sigma_{K,h}^2 = \frac{1}{K^2} \mathbb{E}_{\pi_S} [(A(Z_1) - A(Z_0))^2].$$

From (3.1), we deduce that

$$f_K(Z_1) - f_K(Z_0) = (K - 1)(Z_{1,1} - Z_{0,1}) - [A(Z_1) - A(Z_0)].$$

From the reversibility of the Markov chain $(F_n)_n$, we have

$$\mathbb{E}_{\pi_S} [(A(Z_1) - A(Z_0))(f_K(Z_1) - f_K(Z_0))] = 0.$$

Indeed, $\mathbb{E}_{\pi_S} [(A(Z_1) - A(Z_0))f_K(Z_1)]$ can be rewritten (since $f_K(z)$ depends only on $\mathcal{F}(z)$) as

$$\begin{aligned} \mathbb{E}_{\pi_S} [\Delta A(F_0, F_1)f_K(F_1)] &= \sum_{F \in \overline{S}h_{K,h}} \pi_S(F) \sum_{F' \neq F} p(F, F') \Delta A(F, F') f_K(F') \\ &= \sum_{F \in \overline{S}h_{K,h}} \sum_{F' \neq F} \pi_S(F') p(F', F) \Delta A(F, F') f_K(F') \\ &= - \sum_{F \in \overline{S}h_{K,h}} \sum_{F' \neq F} \pi_S(F') p(F', F) \Delta A(F', F) f_K(F') \\ &= -\mathbb{E}_{\pi_S} [\Delta A(F_0, F_1)f_K(F_0)] \\ &= 0 \end{aligned}$$

from Lemma 3.1. Therefore,

$$\sigma_{K,h}^2 = \frac{1}{K} \mathbb{E}_{\pi_S} [(A(Z_1) - A(Z_0))(Z_{1,1} - Z_{0,1})].$$

From Proposition 2.3 and the expression of the measure π_S , we deduce

$$\sigma_{K,h}^2 = \frac{1}{K} \frac{A_{K,h}}{B_{K,h}},$$

where

$$A_{K,h} = 2 \sum_{\substack{z,z' \in \mathcal{C}_{K,h} \\ z' \in \Gamma^+(z)}} (A(z') - A(z)) \quad \text{and} \quad B_{K,h} = \sum_{F \in \overline{S}h_{K,h}} (\deg_S(F) + 1).$$

Applying Propositions 2.4 and 2.1 combined with (2.1), we get

$$A_{K,h} = 2 \sum_{k=0}^{\lfloor K/2 \rfloor} \binom{K}{2k+1} \binom{K-2k}{g-k} \quad \text{and} \quad B_{K,h} = 2|\mathcal{M}_{K,h}| = 2 \sum_{k=0}^{\lfloor K/2 \rfloor} \binom{K}{2k} \binom{K-2k}{g-k}.$$

3.3. Properties of the variance

We are interested in the properties of the variance $\sigma_{K,h}^2$. Denote by $(S_n)_{n \geq 0}$ the random walk on \mathbb{Z} starting from 0 and moving according to the following rule:

$$\mathbb{P}[S_{n+1} = y | S_n = x] = \frac{1}{3} \quad \text{if } y \in \{x-1, x, x+1\}.$$

Then, for any positive integer K and any $h \in \{0, \dots, K\}$ with $K - h$ even,

$$\mathbb{P}[S_K = h] = \frac{|\mathcal{M}_{K,h}|}{3^K},$$

where $|\mathcal{M}_{K,h}|$ denotes the number of Grand Motzkin paths from $(0,0)$ to (K,h) . The variance $\sigma_{K,h}^2$ can then be rewritten as

$$\sigma_{K,h}^2 = \frac{1}{K} \frac{\mathbb{P}[S_K = h+1] + \mathbb{P}[S_K = h-1]}{\mathbb{P}[S_K = h]}. \tag{3.3}$$

Item (i) directly follows from the local central limit theorem for the random walk $(S_n)_{n \geq 0}$ (see for instance [6, Proposition P4, p. 46]).

Note that from the symmetry of the random walk $(S_n)_n$, we have

$$\sigma_{K,0}^2 = \frac{2}{K} \frac{\mathbb{P}[S_K = 1]}{\mathbb{P}[S_K = 0]}.$$

From the Fourier inversion formula, for any $x \in \mathbb{Z}$, we have

$$\begin{aligned} \mathbb{P}[S_n = x] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\theta x} \mathbb{E}[e^{i\theta S_n}] d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\theta x} \mathbb{E}[e^{i\theta S_1}]^n d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\theta x} \left(\frac{1 + 2 \cos \theta}{3} \right)^n d\theta, \end{aligned}$$

since the increments of the random walk $(S_n)_n$ are independent and identically distributed according to the uniform distribution on the discrete set $\{-1, 0, +1\}$. Therefore, we deduce

$$3^K \mathbb{P}[S_K = 1] = (2\pi)^{-1} \left| \int_{-\pi}^{\pi} (1 + 2 \cos(\theta))^K e^{i\theta} d\theta \right| < (2\pi)^{-1} \int_{-\pi}^{\pi} (1 + 2 \cos(\theta))^K d\theta = 3^K \mathbb{P}(S_K = 0),$$

which yields assertion (ii).

Let us now prove item (iii). Since the increments of the random walk $(S_n)_n$ are symmetric and bounded, from Theorem 2.3.5 (formula (2.23)) in [5], there exists a constant c such that, for any n and x ,

$$\left| \mathbb{P}[S_n = x] - \frac{\sqrt{3}e^{-3x^2/(4n)}}{2\sqrt{\pi n}} \right| \leq \frac{c}{n^{3/2}}. \tag{3.4}$$

Now concerning (iii) it is enough to prove that the limit as m tends to infinity of the sequence

$$u_K := \frac{2 \mathbb{P}[S_K = 0]}{(K + 2)(\mathbb{P}[S_K = 0] - \mathbb{P}[S_K = 1])}$$

is strictly greater than one. From (3.4), we have

$$\begin{aligned} u_K &= \frac{2}{K + 2} \frac{\frac{\sqrt{3}}{\sqrt{4\pi K}} + O(K^{-3/2})}{\frac{\sqrt{3}}{\sqrt{4\pi K}}(1 - e^{-3/4K}) + O(K^{-3/2})} \\ &= \frac{2}{K + 2} \frac{\frac{\sqrt{3}}{\sqrt{4\pi K}} + O(K^{-3/2})}{\frac{3\sqrt{3}}{4\sqrt{4\pi K}^{3/2}} + O(K^{-3/2})} \\ &\sim \frac{8}{3} \quad \text{as } K \rightarrow +\infty. \end{aligned}$$

4. The unconstrained case

In this section we indicate how to adapt the ideas we developed in order to give an alternative proof of the main result of [1]. We now consider the Markov chain $(Z_n^*)_{n \geq 0}$ with values in the state space

$$\mathcal{C}_K = \{z \in \mathbb{Z}^{K+1}; \forall i \in [K], |z_{i+1} - z_i| = 1\},$$

so that \mathcal{C}_K is the union $\cup_h \mathcal{C}_{K,h}$. The difference is that both ends of the walks are allowed to move in different directions. The set of paths $\Gamma(z) = \Gamma^+(z) \sqcup \Gamma^-(z)$ is defined as before.

By Proposition 2.1, Φ^+ is a bijection between

$$\{(z, z') \in \mathcal{C}_K^2; z' \in \Gamma^+(z)\}$$

and $\mathcal{M}_K := \cup_h \mathcal{M}_{K,h}$, which clearly has cardinality 3^K since each step can be chosen independently. Now modify the area of $z \in \mathcal{C}_K$ by setting

$$A^*(z) := \frac{3}{2}z_1 + z_2 + \dots + z_K + \frac{3}{2}z_{K+1} = z_1 + A(z) + z_{K+1}.$$

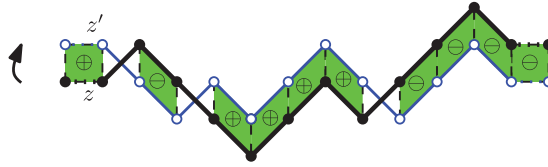


Figure 5. The area $A^*(z') - A^*(z)$ is equal to $1 - 1 + 4 - 4 = 0$.

Extend any path $z \in \mathcal{C}_K$ by unit horizontal steps at both ends: then $A^*(z)$ is the algebraic area below this extended path. We have the following proposition, which is the counterpart of Propositions 2.3 and 2.4 in this unconstrained setting.

Proposition 4.1. For any $z \in \mathcal{C}_K$,

$$\sum_{z' \in \Gamma(z)} (A^*(z') - A^*(z)) = 0.$$

For any K ,

$$\sum_{\substack{z, z' \in \mathcal{C}_K \\ z' \in \Gamma^+(z)}} [A^*(z') - A^*(z)] = 2 \cdot 3^K.$$

Proof Sketch of the proof. The difference $A^*(z') - A^*(z)$ between two paths can be written as the sum of $\text{sign}(z; z', p)$ over all marked paths $(z; z', p)$, where the mark p can now also be one of the two extra horizontal steps; see Figure 5. We need to define \mathcal{I}^* , a modification of the involution \mathcal{I} from Section 2. Given a marked path $(z; z', p)$, its image $(z; z'', q)$ by \mathcal{I}^* is defined as follows. If p is the initial horizontal step, then find the first crossing: z'' is obtained by changing z'' at this step, which becomes the new mark q (if there is no crossing, so that $z' = z^\pm$, then change it to $z'' = z^\mp$ and let q be the final horizontal step). Define \mathcal{I}^* symmetrically when p is the final horizontal step. In all other cases, define \mathcal{I}^* as in \mathcal{I} , except in the cases where one needs to use the special cases of L and R : in this case the mark q will be one of the horizontal steps. It is easily seen that \mathcal{I}^* is bijective and reverses signs, so that the first formula of the proposition is proved.

For the second one, we notice as in Section 2 that the left-hand side can be written as the sum of $\text{sign}(z; z', p)$ over all marked paths satisfying $z' \in \Gamma^+(z)$ and $z'' \in \Gamma^-(z)$. This happens when p is the initial horizontal step for any z, z' with $z' \in \Gamma^+(z)$, and these cases contribute 3^K to the sum. The other possibility is that p is a down-step which has no crossing to its left. Clearly $\text{sign}(z; z', p) = 1$ here also since $z' \in \Gamma^+(z)$, and there are also 3^K such possibilities: indeed, their images by \mathcal{I}^* in this case are exactly the marked paths $(z; z'', q)$ where $z'' \in \Gamma^-(z)$ and q is the initial horizontal step. This completes the proof. \square

Let $f_K^*(z)$ denote the area between a path $z = (z_1, z_2, \dots, z_K, z_{K+1}) \in \mathcal{C}_K$, and the segment $(z_1, z_1 + 1, \dots, z_1 + K + 1, z_1 + K + 2)$. Then, for any $z \in \mathcal{C}_K$ (see Figure 6),

$$f_K^*(z) + A^*(z) = (K + 2)z_1 + \frac{(K + 2)^2}{2}. \tag{4.1}$$

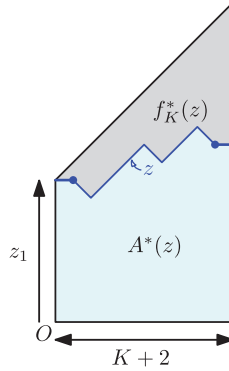


Figure 6. Relation between the first coordinate z_1 and the area under the path.

Since f_K^* is bounded by $(K + 2)^2$, the position of the first walker and the area under the extended path can be asymptotically related as

$$\frac{Z_{[n],1}^*}{\sqrt{n}} = \frac{1}{K + 2} \frac{A^*(Z_{[n]}^*)}{\sqrt{n}} + o(1). \tag{4.2}$$

The first assertion in Proposition 4.1 implies that the sequence $(A^*(Z_n^*))_{n \geq 0}$ is a martingale with respect to the natural filtration of the Markov chain. The main result of [1] is then deduced from the functional central limit theorem for the rescaled martingale $(A^*(Z_n^*)/\sqrt{n})_{n \geq 1}$. The proof is similar to the one given in Section 3 in the constrained case. The asymptotic variance is then rewritten as

$$\sigma_{K,*}^2 = \frac{1}{K + 2} \frac{A_K^*}{B_K^*},$$

where

$$A_K^* = 2 \sum_{\substack{z, z' \in C_K \\ z' \in \Gamma^+(z)}} (A^*(z') - A^*(z)) \quad \text{and} \quad B_K^* = 2|\mathcal{M}_K| = 2 \cdot 3^K.$$

The second assertion in Proposition 4.1 leads to

$$\sigma_{K,*}^2 = \frac{2}{K + 2}.$$

5. Discussion and open problems

(1) The case when the distance h between the extremal coordinates depends on K can also be considered. It is not difficult to see that Theorem 1.1 still holds, with h replaced by $h(K)$ in the variance formula. The asymptotic behaviour of the variance for K large (item (i) in Theorem 1.1) is still true if $h(K) = o(K^{3/4})$. Indeed, relation (3.3) is still valid and local limit theorem 2.3.11 in [5, p. 46] gives the result after elementary computations. The asymptotic behaviour of the variance for $h \gg K^{3/4}$ is not known.

(2) Instead of $\mathcal{C}_{K,h}$ we can consider the set of paths

$$\mathcal{C}_K = \{z \in \mathbb{Z}^{K+1}; \forall i \in [K], |z_{i+1} - z_i| \in \{0, 1\}\}$$

and $(Z_n)_{n \geq 0}$, the Markov chain defined on \mathcal{C}_K corresponding to $K + 1$ simple random walks on \mathbb{Z} under the shape constraint. In [1], the set of paths

$$\mathcal{D}_K = \{z \in \mathbb{Z}^{K+1}; \forall i \in [K], |z_{i+1} - z_i| = 1\}$$

and the corresponding Markov chain $(Z_n^{(K)})_{n \geq 0}$ were studied. Whatever the distribution of $Z_0^{(K)}$, the sequence of random variables $(Z_{n,1}^{(K)}/\sqrt{n})_n$ converges to a centred Gaussian law with variance $2/(K + 2)$. By noting that, given the number of zeros N_0 of the initial random variable Z_0 , the position of the first walkers $(Z_{n,1})_n$ and $(Z_{n,1}^{(K-N_0)})$ are identically distributed (with the following convention: if $N_0 = K$, the first random walker evolves as the simple symmetric random walk on \mathbb{Z}), it follows that the sequence of random variables $(Z_{n,1}/\sqrt{n})_n$ converges in law to a mixture of Gaussian distributions with variance equal to

$$\sigma_K^2 = 2 \sum_{l=0}^K \frac{\mathbb{P}[N_0 = l]}{K + 2 - l}.$$

Note that whatever the distribution of the random variable N_0 , the variance is still greater than that obtained in the unconstrained case considered in [1] and is rational when Z_0 is uniformly distributed on \mathcal{C}_K .

(3) The generalization of Theorem 1.1 to more general sets of paths does not seem to be obvious. Except in the case of [1], where our proof can be adapted, the construction of a convenient martingale from which the computation of the variance can be deduced is far from clear. For instance, consider the set of Bernoulli bridges with length K with the additional constraint that L of the $K + 1$ random walkers stay at the same height at each step. A central limit theorem for the first random walker should also hold. We conjecture that the variance should be increasing in L .

Acknowledgements

We are grateful to James Norris and Serge Cohen for stimulating discussions.

References

- [1] Boissard, E., Cohen, S., Espinasse, T. and Norris, J. (2014) Diffusivity of a random walk on random walks. *Random Struct. Alg.*, to appear. [arXiv:1210.4745](https://arxiv.org/abs/1210.4745)
- [2] Durrett, R. (1991) *Probability: Theory and Examples*, Wadsworth & Brooks/Cole.
- [3] Gordin, M. I. (1969) The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188** 739–741.
- [4] Häggström, O. and Rosenthal, J. S. (2007) On variance conditions for Markov chain CLTs. *Electron. Comm. Probab.* **12** 454–464.
- [5] Lawler, G. and Limic, V. (2010) *Random Walk: A Modern Introduction*, Cambridge University Press.
- [6] Spitzer, F. (1976) *Principles of Random Walks*, second edition, Springer.