Experimental analysis of lattice walks

Anthony Zaleski*

February 26, 2015

Abstract

In his book *An Introduction to Probability Theory and Its Application*, Feller discusses statistics corresponding to sequences of coin tosses, with a dollar being won or lost depending on the outcome of each toss. This is equivalent to analyzing walks in the plane with each step being one unit up or right.

One statistic corresponding to a walk is the "number of losing times," i.e., the number of times the player has negative money, or the number of times the 2-D walk is below the line $y = x$. Two main results for this random variable are that it follows the "discrete arcsine law" when the sample space is walks of length n, and it is described by the Chung-Feller generating function when the sample space is walks to (n, n).

In his paper "Fully AUTOMATED computerized redux of Feller's (v.1) Ch. III (and much more!)" (`http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/feller.html`) and the accompanying Maple package, Zeilberger outlines a way to obtain Feller's results with the help of a computer. In fact, Zeilberger's Maple package computes the "grand generating function," which, in a single blow, captures information about *all* of the walk statistics discussed by Feller.

In this paper, we continue to investigate walks using computer methods. However, we shall introduce an approach different from that of Zeilberger, who used computer algebra is used to exactly compute the weight enumerator over *all* walks—an *infinite* sum expressed as

*Department of Mathematics, Rutgers University (New Brunswick), 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA.

an algebraic function. Our procedures input a *numeric* value of n and use dynamic programming methods to find the weight enumerator over all walks of length n—a *finite* polynomial. Then, by generating these polynomials for many values of n, we are able to conjecture behavior of the moments of certain statistics as the walk length tends to ∞.

The advantage of this method is that it is easily applied to more general problems not amenable to an analytic approach. For example, we use it to analyze walks in three dimensions and walks where more general steps (e.g., diagonal steps) are allowed.

1 Introduction

1.1 Coin tossing and walks in the plane

Suppose a gambler tosses a coin finitely many times, winning a dollar whenever heads comes up and losing a dollar when tails appears. The evolution of the game can be described by a string $w = w_1 w_2 \cdots w_n$, where $w_i \in \{-1, 1\}$ describes the outcome of the ith coin toss.

By making the association $-1 = r$ (a right step) and $1 = u$ (an up step), we can also interpret a string w as a walk in $\mathbb{N}^2$ starting from $(0,0)$. For example, the string ur represents a game in which a dollar is won and then lost. Equivalently, ur is a walk from $(0,0) \to (1,0) \to (1,1)$. We shall use the terms "game," "string," and "walk" interchangeably.

Definition 1.1. *Let W be the set of all such walks. We have the following walk statistics (functions from $W \to \mathbb{N}$):*

- *the length (number of steps), $l(w)$;*

- *the number of losing times (points where the walk is below $y = x$):*

$$a_1(w) := \left| \left\{ i : \sum_{j=1}^{i} w_j < 0 \ or \ \left(\sum_{j=1}^{i} w_j = 0 \ and \ \sum_{j=1}^{i-1} w_j < 0 \right) \right\} \right|;$$

- *the number of break-even times (points on $y = x$):*

$$a_2(w) := \left| \left\{ i : \sum_{j=1}^{i} w_i = 0 \right\} \right|;$$

- *the last break-even time:*

$$a_3(w) := \max\left\{ i : \sum_{j=1}^{i} w_j = 0 \text{ and } \sum_{j=1}^{i} w_j > 0 \text{ for } r > i \right\};$$

- *and the number of sign-changes (points where the walk crosses $y = x$):*

$$a_4(w) := \left| \left\{ i : \sum_{j=1}^{i-1} w_j \cdot \sum_{j=1}^{i+1} w_j < 0 \right\} \right|.$$

Further, we define W_n to be the set of n-step walks in W, and $W_{n,n}$ to be the walks to (n, n).

1.2   Old results

Theorem III.4.1 in [1] states that there are $\binom{2k}{k}\binom{2n-2k}{n-k}$ walks $w \in W_n$ satisfying $a_1(w) = 2k$. For n large and fixed, the distribution of a_1 resembles (modulo scaling) $1/\sqrt{x(1-x)}$, so it is sometimes called the *discrete arcsine* distribution. It is u-shaped, meaning that, surprisingly (or not surprisingly, if you believe in luck), most walks are either winning for most of their duration or losing on the majority of flips.

Theorem III.9 in [1], the Chung-Feller theorem, says that the number of walks to (n, n) with $2k$ losing times is given by a Catalan number, and is independent of k. In terms of generating functions:

$$\sum_{w \in W_{n,n}} t^{a_1}(w) = \frac{1}{n-1}\binom{2n}{n}\sum_{k=0}^{n} t^{2k}.$$

Finally, in [2], Zeilberger uses Maple to evaluate the "grand generating function"

$$\sum_{w \in W} z^{l(w)} t_1^{a_1(w)} t_2^{a_2(w)} t_3^{a_3(w)} t_4^{a_4(w)}$$

as a (very messy) algebraic function of z, t_1, t_2, t_3, t_4.

These results are illuminating for this problem, but they are gotten through ad hoc methods. So, for example, it is not obvious how to derive analogous results for walks in higher dimensions, or walks where non-standard steps are allowed.

3

Here, we shall give an alternative method to analyze the statistics of a very general class of walks and approximate the long-run behavior of their moments. But first, let us see if we can discover a few more exact results using the analytical approach.

2 Exact results

2.1 Moments of up-right walks to (n, n)

Suppose we uniformly randomly pick a walk $w \in W_{n,n}$. Then we can think of $a_1(w)$ as a random variable. For each n, $a_1|W_{n,n}$ ($| =$ "with sample space") has a certain distribution, so it is natural to wonder about the limiting distribution as $n \to \infty$. For example, is it asymptotically normal?

Recall that the moments of a random variable are a fingerprint of its distribution; for example, the (standardized central) moments $0, 1, 0, 3, 0, 15, 0, 105, \ldots$ uniquely determine the normal distribution.

In this case of $a_1|W_{n,n}$, we can find the moments in terms of n, which is not surprising since a_1 is essentially uniform by the Chung-Feller rule.

Using the procedure `ChungFeller` in [2], we find

$$\sum_{w \in W_{n,n}, \, n \in \mathbb{N}} z^n t^{a_1(w)} = \frac{2}{\sqrt{-4\,z+1} + \sqrt{-4\,zt^2+1}}.$$

Now we use `convert(%,FPS,z)` to convert this function to a formal power series in z. By looking at the coefficient of z^n, we obtain the generating function

$$F_n(t) = \sum_{w \in W_{n,n}} t^{a_1(w)},$$

as a function of n! To see it for yourself, use `ChungFellerGF(t,n)`.

Now it is easy to compute the moments as functions of n, which is done in `ChungFellerMoment`. You can easily verify the following:

Proposition 2.1. *The number of losing times of a walk chosen uniformly randomly from $W_{n,n}$ has mean n and variance $n^2/3 + 2n/3$, and its third through tenth standardized moments about the mean approach $0, 9/5, 0, 27/7, 0, 9, 0, 243/11$ as $n \to \infty$.*

Analogously, for a_2, the number of visits to the diagonal $y = x$, we have

$$F(z, t) := \sum_{w \in W_{n,n}, n \in \mathbb{N}} z^n t^{a_2(w)} = \frac{1}{t\sqrt{-4z + 1} - t + 1}.$$

Unfortunately, Maple cannot convert this to a formal power series in z. However, $F_t(z, 1)$ *is* convertible to a formal power series, so we can compute $\mathbb{E}[a_2 | W_{n,n}] = [z^n] F_t(z, 1)/\binom{2n}{n}$, as a function of n. In a similar way, we can find higher moments: the idea is to repeatedly apply the operator $t\partial_t$, substitute $t = 1$, and *then* expand as a formal power series in z. (Some conversions are necessary to get the central moments.)

The moments of a_2 are surprisingly complicated in comparison with those of a_1:

Proposition 2.2. *The number of visits to $y = x$ of a walk chosen uniformly randomly from $W_{n,n}$ has mean and variance*

$$\frac{-(2n)! + 4^n (n!)^2}{(2n)!}, \quad -\frac{16^n (n!)^4 + 4^n (n!)^2 (2n)! - 4n ((2n)!)^2 - 2 ((2n)!)^2}{((2n)!)^2},$$

and its third through fifth standardized moments about the mean approach

$$2\frac{\sqrt{\pi}(\pi - 3)}{(-\pi + 4)^{3/2}}, \quad -\frac{3\pi^2 - 32}{\pi^2 - 8\pi + 16}, \quad 4\frac{\sqrt{\pi}(\pi^2 + 5\pi - 25)}{(-\pi + 4)^{5/2}}$$

as $n \to \infty$.

2.2 Forward King walks

Now we examine another special set of walks:

Definition 2.3. *Let $K_{n,n}$ be the set of walks from $(0, 0) \to (n, n)$ with steps in $\{r, u, d\} = \{(1, 0), (0, 1), (1, 1)\}$. Let $K := \bigcup_{n \in \mathbb{N}} K_{n,n}$. For $w \in K$, let $n(w)$ be the n such that $w \in K_{n,n}$.*

Think of K as the set of journeys possible for a forward-marching King that end on the line $y = x$. Each move, we take a step from $\{r, u, d\}$ (right, up, or diagonal).

For a set $E \subset K$, define the generating function

$$F_E(z, t) := \sum_{w \in E} z^{n(w)} t^{a_1(w)}.$$

We we shall find an algebraic expression for $F_K(z,t)$. The idea is to convert facts describing walks in K to equations involving generating functions.

Definition 2.4.

- *Juxtaposition of two sets A, B of walks denotes concatenation:*

$$AB := \{w_1 w_2 : w_1 \in A, w_2 \in B\}.$$

  If A or B is a singleton, we drop the braces: e.g., $aB := \{a\}B$ for a walk a.

- *The* Kleene star *of a set of walks is its closure under concatenation:*

$$E^* := E \cup EE \cup \cdots = \{s_1 s_2 \cdots s_k : k \in \mathbb{N}, s_k \in E\}.$$

- *We define the* star *of a generating function F to be*

$$F^* := 1 + F + F^2 + \cdots = \frac{1}{1 - F}.$$

Now, let N be the negative walks in K, i.e., walks satisfying $y < x$, save for the first and last points. Let Φ be the nonpositive walks, i.e., walks in $y \leq x$. Any negative walk is a right step followed by a nonpositive walk followed by an up step: $N = r\Phi u$. Note that *every* point of $w \in \Phi$ is counted as a losing time in rwu, so defining

$$\tilde{F}_\Phi := \sum_{w \in \Phi} z^{n(w)} t^{l(w)},$$

we have

$$F_N = zt^2 \tilde{F}_\Phi. \tag{2.1}$$

Next, any nonpositive walk consists of diagonal steps and negative walks. So $\Phi = d^*(Nd^*)^*$, and

$$\tilde{F}_\Phi = (zt)^*(F_N(zt)^*)^*. \tag{2.2}$$

It is child's play for Maple to solve (2.2) and (2.2) for $F_N(z,t)$. Now let P be the set of positive walks, i.e., walks in $y > x$, except for the endpoints. Positive walks are simply negative walks reflected about $y = x$, so $F_P = F_N(z,1)$ (all positive walks have zero losing times).

Finally, any forward King walk consists of diagonal steps, negative walks, and positive walks: $K = d^*(Nd^* \cup Pd^*)^*$. In terms of generating functions,

$$F_K = z^*(F_N z^* + F_P z^*)^*,$$

and we are finished! We have $F_K(z,t)$ as an algebraic expression. Of course, it is rather messy, so we do not record it here. To see it for yourself, use `ForwardKingGF(z,t)` in the Maple package. Unfortunately, F_K is too complex to be amenable to either of the moment-finding methods discussed previously. However, we should not lose hope...

3 Numerically analyzing moment asymptotics

3.1 Recursive enumeration of walks

We started with steps in $\{(1,0),(0,1)\}$. Then we added the diagonal step $(1,1)$. Now let us take the affair even further.

Definition 3.1. *Given $S \subset \mathbb{N}^2$, let W^S be the set of walks from $(0,0)$ with steps in S. For $(a,b) \in \mathbb{N}^2$, let $W^S_{a,b}$ contain walks of W^S ending at (a,b).*

In the $S = \{(1,0),(0,1)\}$ case, we were able to calculate the moments of $a_1|W^S_{n,n}$ in terms of n. We cannot expect to do this in general. Indeed, even in the (still very symmetric) case $S = \{(1,0),(0,1),(1,1)\}$, we could not find nice expressions for the moments. For even wilder choices of S, who knows if closed form generating functions over W^S even exist?

However, we can *fix* $(a,b) \in \mathbb{N}^2$ and focus on the *finite* set of walks $W^S_{a,b}$. Then the generating function

$$F_{a,b}(t) := \sum_{w \in W^S_{a,b}} t^{a_1}$$

is a *finite* polynomial in t, with easily computable moments. Further, given fixed S and (a,b), we can make use of the fact that

$$W^S_{a,b} = \bigcup_{s \in S} W_{(a,b)-s}\{s\}$$

to compute $F_{a,b}(t)$ with a recursive procedure; this is done in `F2G`. So for each (a,b), the moments of $a_1|W^S_{a,b}$ are easily computable.

7

3.2 Asymptotic storybooks

The procedure `ChungFellerBook2D(S,M,K1,K2)` uses `F2G` to compute the expectation, variance, and central standardized moments three through M of $a_1 | W_{n,n}^{S'}$ for $S' \subset S$, $n = K1, \ldots, K2$. It uses this data to guess the asymptotic behavior of the moments as functions of n. We use the ansatzes Cn for expectation, Cn^2, and C for the third and higher central standardized moments.

So, for each $S' \subset S$, a theorem about the asymptotic behavior of walks with steps in S' is generated (step sets producing trivial theorems are automatically excluded).

Of course, we must add the disclaimer that these "theorems" are merely numerical approximations to the asymptotic behavior of the moments. To be extra safe, the procedure `ChungFellerBook2DSafe` runs `ChungFellerBook2D` twice, with different n-ranges. For each theorem it computes the constants twice; then it only keeps the agreeing digits.

In the case $S = \{(1,0),(0,1)\}$, where we do know the moments as functions of n, we can confirm that `ChungFellerBook2DSafe` gives good results.

The table below summarizes the output of
`ChungFellerBook2DSafe({[1,0],[0,1],[1,1],[2,0],[0,2]},6,100,110,`
`190,200);`. The zeroth column is the set of allowed steps, where for brevity $ij := [i,j]$. Columns 1-6 are the expectation, variance, and limits of the third through sixth standardized central moments. Note that by Proposition 2.1, the exact values of the first row are $n, n^2/3, 0, 1.8, 0, 27/7 \approx .38571$.

8

Steps	1	2	3	4	5	6
$\{01, 10\}$	$1.0000n$	$0.3n^2$	0.0000	1.800	0.0000	3.86
$\{01, 20\}$	$0.38n$	$0.1n^2$	0.0	0.900	-0.1	1.93
$\{02, 20\}$	$0.2500n$	$0.043n^2$	0.0000	0.900	0.0000	1.93
$\{01, 02, 10\}$	$0.9n$	$0.27n^2$	0.0	1.8023	-0.03	3.87
$\{01, 02, 20\}$	$0.33n$	$0.07n^2$	-0.02	0.90	-0.1	$2.$
$\{01, 10, 11\}$	$0.8n$	$0.2n^2$	0.0	1.8	$0.$	3.9
$\{01, 11, 20\}$	$0.666n$	$0.15n^2$	0.001	1.80	0.01	3.9
$\{02, 11, 20\}$	$0.5n$	$0.08n^2$	0.0	1.80	$0.$	3.9
$\{01, 02, 10, 11\}$	$0.81n$	$0.22n^2$	$0.$	1.80	0.0	3.9
$\{01, 02, 10, 20\}$	$0.80n$	$0.21n^2$	-0.01	1.804	$0.$	$4.$
$\{01, 02, 11, 20\}$	$0.6n$	$0.1n^2$	-0.01	1.80	-0.1	3.9
$\{01, 10, 11, 20\}$	$0.81n$	$0.22n^2$	$0.$	1.80	0.03	3.89
$\{01, 02, 10, 11, 20\}$	$0.75n$	$0.19n^2$	-0.004	1.8	-0.011	$4.$

3.3 Walks in higher dimensions

This method easily generalizes to three or more dimensions. If we consider walks in $\mathbb{N}^3$, then we have seven statistics to keep track of the number of times the walk visits the regions $x < y < z$, $x < z < y$, $y < x < z$, $y < z < x$, $z < x < y$, $z < y < x$, and "none of the above." The corresponding generating function (over walks to a fixed point in $\mathbb{N}^3$) is computed in `F3G`. Not surprisingly, this procedure is significantly slower than `F2G`.

4 Conclusion

Many areas still need to be explored. For example, we have focused mainly on the number of losing times, a_1. But the method of Section 3 could also be applied to the other statistics in Definition 1.1. Also, there is much to be done with walks in higher dimensions. Our main goal in this paper was to illustrate by example that experimental mathematics can give us insights into problems where exact analysis is difficult. Hopefully the results here are just the beginning. We encourage you to experiment with the Maple package and make discoveries of your own!

5 Using the Maple package

The Maple package accompanying this paper may be found at `http://www.math.rutgers.edu/~az202/Z/Feller.txt`. To use it, place it in the working directory and execute `read('Feller.txt');`. To see the main procedures, execute `Help();`. For help on a specific procedure, use `Help(procedure_name);`. Happy exploring!

Acknowledgement

References

[1] William Feller. *An Introduction to Probability Theory and Its Application*, volume 1. John Wiley and sons, 3 edition, 1968.

[2] Doron Zeilberger. Fully AUTOMATED computerized redux of Feller's (v.1) Ch. III (and much more!). *Personal Journal of Shalosh B. Ekhad and Doron Zeilberger*, November 2006. `http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/feller.html`.