An Introduction to a Small Variety of Simple Random Walks

Austin Barton[1]

[1]MATH 3235, Georgia Institute of Technology

December 14, 2022

A Brief Introduction

This paper primarily explores and discusses simple random walks and their intersection between enumerative combinatorics and probability. We will develop the basic framework to understanding simple random walks in a common manner utilizing textbooks such as *Probability: An Introduction* by Grimmett and Welsh and *Probability and measure* by Billingsley. Then, we will shift to enumerative combinatorics by developing basic ideas on Dyck paths using generating functions and conclude by connecting them to simple random walks as a stochastic process.

Random walks are sequences of non-deterministic events. They are ubiquitous in nature and fundamental in many subjects. Random walks describe many different processes and offer a lot of information on various scenarios when studied. Random walks surface in many areas such as combinatorics and graph theory, polymer physics, estimating the size of the Web, financial economics, Fermi estimation, ergodic theory, and many more.

We assume some familiarity and elementary understanding of what a path is in mathematics. Simple random walks in this paper, and in many case, will be visualized and studied as paths with discrete steps in the Cartesian plane where the abscissa (horizontal component of a coordinate) represents the time and the ordinate (vertical component) represents the state of the system.

Random Walks

1 Ideas, Definitions, and some Theorems

What follows is a series of some basic groundwork necessary to begin to study random walks. Much of it is directly sourced from Grimmett and Walsh's *Probability: An Introduction*. However, we adapt it slightly using Hoel's *Introduction to Stoachastic Processes* in order to introduce more abstract notions of systems and state spaces which we believe would be useful in future studies of stochastic processes.

Definition 1.1. A **stochastic process** is generally defined as any collection of random variables $X(t)$, $t \in T$, defined on a common probability space, where $T \subseteq \mathbb{R}$. (Hoel, 1972)

That is, a stochastic process is generally any collection of random variables and they are typically used to model non-deterministic systems and phenomena. In our case, and many others, we will study discrete systems and thus $T \subseteq \mathbb{Z}$. It is important to note that the set of parameters on which our collection of random variables is defined, T, need not be discrete. T will be referred

throughout this paper, and many others, as the time. Therefore, t will only take values $0, 1, 2, \ldots$ and thus we let $T = \mathbb{N} \cup \{0\}$

Definition 1.2. A state is a certain configuration of a system. A **state space** is the set of all possible configurations/states of a system. We will denote $\mathscr{I}$ to be the state space of the stochastic process for which we are interested in. For this paper, the state space is assumed to be *discrete* and $\mathscr{I} = \mathbb{Z}$.

Definition 1.3. A **random walk** is a stochastic process that describes the path that consists of a succession of random steps on some mathematical space.

We will observe the state of our system at times $0, 1, 2, \ldots$ and at each of these points in time, the system will have one of the states $\ldots, -2, -1, 0, 1, 2, \ldots$. Let $0 < p < 1$ be the probability that the system "jumps" from its current state to its current state plus 1. Let $q = 1 - p$ be the probability that it "jumps" from its current state to its current state plus -1. If S_n denotes the state of the system at time n, then

$$S_{n+1} = \begin{cases} S_n + 1 & \text{with probability } p, \\ S_n - 1 & \text{with probability } q, \end{cases} \tag{1.1}$$

and we assume that the direction of each jump is independent of all earlier jumps. Therefore,

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n \qquad \text{for } n \in T \tag{1.2}$$

where S_0 is the initial state of the system and $X_1, X_2, \ldots$ are independent and identically distributed random variables each taking either value -1 or 1 with probabilities $\mathbb{P}(X_n = 1) = p$ and $\mathbb{P}(X_n = -1) = 1 - p = q$.

Definition 1.4. We call the process $S_0, S_1, \ldots$ a **simple random walk**. We call it *symmetric* if $p = q = \frac{1}{2}$ and *asymmetric* otherwise.

Theorem 1.5. Let S_n be a simple random walk where $S_0 = 0$. $\mathbb{P}(S_n = 0)$ is the probability that S_n is at its starting point at time n. Then

$$\mathbb{P}(S_n = 0) = \begin{cases} 0, & \text{if } n \text{ is odd} \\ \binom{n}{n/2} p^{n/2} q^{n/2}, & \text{if } n \text{ is even} \end{cases} \tag{1.3}$$

where p, q are as previously defined.

Proof. For $S_n = 0$, where $S_0 = 0$, we need that in the simple random walk $S_n = X_1 + \cdots + X_n$, there are equally as many X_i, for $i \in \{1, \ldots, n\}$, that take the value 1 as there are X_i that take the value -1. Thus, the total number of steps is $2m$ where m is the number of X_i that take the value 1. Since n must be even in order for $S_n = 0$, the probability that $S_n = 0$ for an odd number of steps n is $\mathbb{P}(S_n = 0) = 0$.

Suppose now that n is even and take $n = 2m$ where m is the number of X_i that take on the value 1. If $2m$ is the total number of steps of the simple random walk and we know that exactly m of those $2m$ steps take on the value of 1, then there are exactly $\binom{2m}{m}$ distinct combinations of X_i in the simple random walk where exactly m of them are $X_i = 1$.

Note that each X_i are identical independent random variables. So, for any one simple random walk of n steps where $S_n = 0$, the probability that m X_i's take on the value 1 and m X_i's take on value -1 is

$$\mathbb{P}(X_i = 1)^m \mathbb{P}(X_i = -1)^m = p^m q^m$$

where i is any $i \in \{1, \ldots, n\}$. That is, for each simple random walk where $S_n = 0$, its corresponding probability of occurring is $p^m q^m$. Since we showed that there are exactly $\binom{2m}{m}$ of these, then our probability that $S_n = 0$ is

$$\mathbb{P}(S_n = 0) = \binom{2m}{m} p^m q^m$$

where $n = 2m$. (Grimmett and Walsh, 2017)

Note that if we assume equiprobability in our simple random walk from Theorem 2.1, if n, then,

$$\mathbb{P}(S_n = 0) = \frac{\text{total number of paths with n steps that end at 0}}{\text{total number of paths with n steps}} = \frac{\binom{n}{n/2}}{2^n} = \frac{\binom{n}{n/2}}{2^{n/2} 2^{n/2}}$$

where 2^n is the enumeration of all paths of length n. This is exactly the result from Theorem 1.5 with parameter $p = 1/2$. $\qquad\square$

Definition 1.6. We call a random walk **recurrent** if it revisits the starting point infinitely often with probability one. In other words, if it is bound to revisit the starting position at some point in time. Otherwise, we call a random walk **transient**. (Grimmett and Walsh, 2017)

Definition 1.7. An **absorbing barrier** is state of the system in which if the system reaches that state at any time, then the process terminates. The **absorption probability** is the probability of the random walk reaching an absorbing barrier.

In the following section, we will study a specific scenario involving the definitions 1.6 and 1.7 in regard to simple random walks.

2   Gambler's Ruin

The gambler's ruin problem is a statistical scenario centered around conditional probabilities and expected outcomes. The problem was around long before random walks were formally discovered/studied (the earliest known mention of the problem being in 1656). This section will cover a brief introduction to the idea and is not meant to be a rigorous treatment. Rather, it serves to showcase simple random walks themselves. We will pose the key questions behind the gambler's ruin, develop the basic mathematical framework using what we discussed in Section 1, and go over a brief derivation of the Gambler's Ruin theorem with an emphasis on intuition in lieu of a proof.

Let $X_1, X_2, \ldots$ be i.i.d. random variables on $(\Omega, , \mathbb{P})$ as defined in (1.2). Suppose that a gambler enters a casino with an initial capital of a and has the stubborn strategy of betting until either they achieve their goal $c > a$ or they are broke and they hove no other choice but to leave the casino. There are two main questions we concern ourselves with. (Billingsley, 1995)

Question: What is the probability of ruin? That is, the probability that they go broke.

Question: What is the probability they reach their goal? That is, the probability they increase their capital from a to c.

Let $S_n = X_1 + \cdots + X_n$, where $S_0 = 0$. The gambler's fortune after n plays or steps of the simple random walk is $a + S_n$. The event

$$A_{a,n} = [a + S_n = c] \cap \bigcap_{k=1}^{n-1} [0 < a + S_k < c] \tag{2.1}$$

represents the success of the gambler after n plays. If $n > c - a$, then there are more ways than one for the gambler to reach success. If $n = c - a$, then the gambler must win every play of the game in order to reach success. And if $n < c - a$, then there aren't enough plays to reach success in that case and $A_{a,n} = \emptyset$. Additionally,

$$B_{a,n} = [a + S_n = 0 \cap \bigcap_{k=1}^{n-1} [0 < a + S_k < c]] \tag{2.2}$$

represents the ruin of the gambler after n plays.

If $s_c(a)$ denotes the probability of reaching success, then,

$$s_c(a) = \mathbb{P}(\bigcup_{n=1}^{\infty} A_{a,n}) = \sum_{n=1}^{\infty} \mathbb{P}(A_{a,n}) \tag{2.3}$$

for $0 < a < c$. Note that it splits since each $A_{a,n}$ is disjoint from one another. Similarly, if $s_0(a)$ denotes the probability of ruin, then,

$$s_0(a) = \mathbb{P}(\bigcup_{n=1}^{\infty} B_{a,n}) = \sum_{n=1}^{\infty} \mathbb{P}(B_{a,n}) \tag{2.4}$$

c and 0 are both absorbing barriers on S_n. $s_c(a)$ is the absorption probability of c. $s_0(a)$ is the absorption probability of 0.

From (Billingsley, 1976) we an intuitive argument for a recurrence relation on the absorption probability $s_c(a)$. Note that X_i are a sequence of i.i.d. random variables in S_n. So, by independence, the sequence $X_2, X_3, \ldots$ is probabilistically identical to $X_1, X_2, \ldots$. Thus, the chance of success for a gambler with initial fortune a must be the chance of winning the first wager times the chance of

success for an initial fortune $a+1$, plus the chance of losing the first wager times the chance of success for an initial fortune $a-1$ (this is essentially the partition theorem). Therefore,

$$s_c(a) = ps_c(a+1) + qs_c(a-1) \tag{2.5}$$

This is a second order linear homogeneous difference equation subject to boundary conditions $s_c(0) = 0$ and $s_c(c) = 1$. That is, if the gambler's initial capital is 0 then there is 0 chance of success and if their initial capital is c then there is chance of success is 1. In either case, we wonder what was going through their heads when they walked into the casino with this goal. We can solve (2.5) obtaining,

$$s_c(a) = \begin{cases} A + B(\frac{q}{p})^a & \text{if } p \neq q \\ A + Ba & \text{if } p = q = \frac{1}{2} \end{cases} \tag{2.6}$$

Using boundary conditions, we find that $A = \frac{1}{1-(q/p)^c}$ and $B = \frac{1}{(q/p)^c-1}$. Thus, after substituting and simplifying for A and B, we have obtained,

Theorem 2.1. (**Gambler's Ruin**) Fix c and let $n \geq 1$ and $0 < a < c$. Consider the simple random walk S_n on the $\{0, 1, 2, \ldots, c\}$ with absorbing barriers on 0 and c. Then, the probability that the walk is absorbed at c is given by

$$s_c(a) = \begin{cases} \frac{(q/p)^a-1}{(q/p)^c-1} & \text{if } p \neq q \\ a/c & \text{if } p = q = \frac{1}{2} \end{cases} \tag{2.7}$$

Equiprobable Paths

3   Enumerating Paths

3.1   Lattice Paths

We assume the reader has some familiarity with Lattice paths commonly seen in combinatorics. Recall that the total number of distinct Lattice paths from a point (a,b) to a point (p,q) is

$$\binom{(p-a)+(q-b)}{p-a} \tag{3.1}$$

We will study "diagonal" Lattice paths as diagonal steps in the first and fourth quadrants of the Cartesian plane where up corresponds to the step $(1,1)$ and down corresponds to the step $(1,-1)$.

Let u be the letter that denotes a rise, the step $(1,1)$, and d be the letter that denotes a fall, the step $(1,-1)$. Each Lattice path is encoded by a word consisting of letters from the alphabet $\{u,d\}$. We denote the lattice word as γ with t number of steps. That is, $\gamma = \gamma_1\gamma_2\ldots\gamma_t \in \{u,d\}^*$.

Let t denote the total number of steps. Let m denote the total number of ups, $(1,1)$, and n denote the total number of downs $(1,-1)$. The total number of steps is $t = m+n$. Then the total number of Lattice paths between any two points is the total number of distinct combinations of u and d in the Lattice path's corresponding word γ, which contains exactly m u's and n d's. This is,

$$\binom{t}{m} = \binom{t}{n}$$

When given the coordinates that the Lattice path ends at, the difference between the start and end abscissa corresponds to t, since each step contains exactly one horizontal component. The difference between the start and end ordinate tells us the difference between the up steps and down steps.

That is, if we are enumerating all the diagonal Lattice paths from (a,b) to (p,q), then

$$t = p-a \qquad\qquad q-b = m-n \tag{3.2}$$

and

$$m = \left\lfloor \frac{t}{2} \right\rfloor + (q-b) = t-n \qquad\qquad n = \left\lfloor \frac{t}{2} \right\rfloor + (b-q) = t-m \tag{3.3}$$

and thus, the total number of distinct Lattice paths from (a,b) to (p,q) is

$$\binom{t}{m} = \binom{t}{n} \tag{3.4}$$

where t,m,n are as previously defined. What is important about how t,m,n are defined is that if we are given the start and end coordinates, then we can find the number of distinct Lattice paths between them.

Typically it is simpler to just transform the start and end coordinates to where the start coordinate is the origin $(0,0)$.

3.2 Dyck Paths: Definitions and Ideas

A **Dyck path** is a path in the first quadrant, $\mathbb{N} \cup \{0\} \times \mathbb{N} \cup \{0\}$, which begins at the origin $(0,0)$ and ends at $(2n, 0)$ where $n \in \mathbb{N} \cup \{0\}$. It consists of steps $(1,1)$ called *rises* and steps $(1,-1)$ called *falls*. n denotes the semi-length of the path. The trivial Dyck path is called the empty Dyck path where $n = 0$. We denote the set of all Dyck paths of semi-length n as $\mathbf{D_n}$. We may refer to Dyck paths of semi-length n as *n*-Dyck paths.

As seen Section 4.1, we let u be the letter that denotes a rise and d be the letter that denotes a fall. Each Dyck path is encoded by a word consisting of letters from the alphabet $\{u, d\}$. That is, for every Dyck path, there is a corresponding word $\alpha = \alpha_1 \alpha_2 \ldots \alpha_{2n} \in \{u, d\}^*$.

A step of a Dyck path with ordinates, with no distinction in order, $k-1$ and k ($k \geq 1$) is said to be at *level* k. A point of a Dyck path with ordinate k is said to be at level k.

A *peak* is an occurrence of the substring ud, a *valley* is an occurrence of du. By the level\height of a peak we mean the level of the intersection point of its two steps. There are other specific occurrences that can be studied such as valleys, du, doublerises, uu, and etc.

By a *return* step we mean a fall at level 1.

If α and β are Dyck paths, then we define

$\alpha\beta$ as the concatenation of α and β.

$\hat{\alpha} := u\alpha d$ as the elevation of α.

(Note that $\hat{\alpha}$ has semi-length of $n \geq 1$ for any Dyck path α)

The set $\hat{\mathbf{D}}_\mathbf{n}$ is defined as

$$\hat{\mathbf{D}}_\mathbf{n} = \{\hat{\alpha} = u\alpha d : \alpha \in \mathbf{D_n}\} \tag{3.5}$$

If $\mathbf{A}$ and $\mathbf{B}$ are finite sets of Dyck paths then $\mathbf{AB}$ is defined as

$$\mathbf{AB} = \{\alpha\beta : \alpha \in \mathbf{A}, \beta \in \mathbf{B}\} \tag{3.6}$$

(Deutsch, 1999)

Example 3.1. Consider a Dyck path of semi-length n. Let α be its corresponding word. Let the substring $ud \in \alpha$. Then, we know that ud is a peak within α. Further, suppose we know that u is a rise from ordinate $k-1$ to k and therefore u is a rise of level k. Then d must be a fall from ordinate k to $k-1$ and therefore is a fall of level k. From our definition of levels of peaks, it follows that ud is a peak of level k.
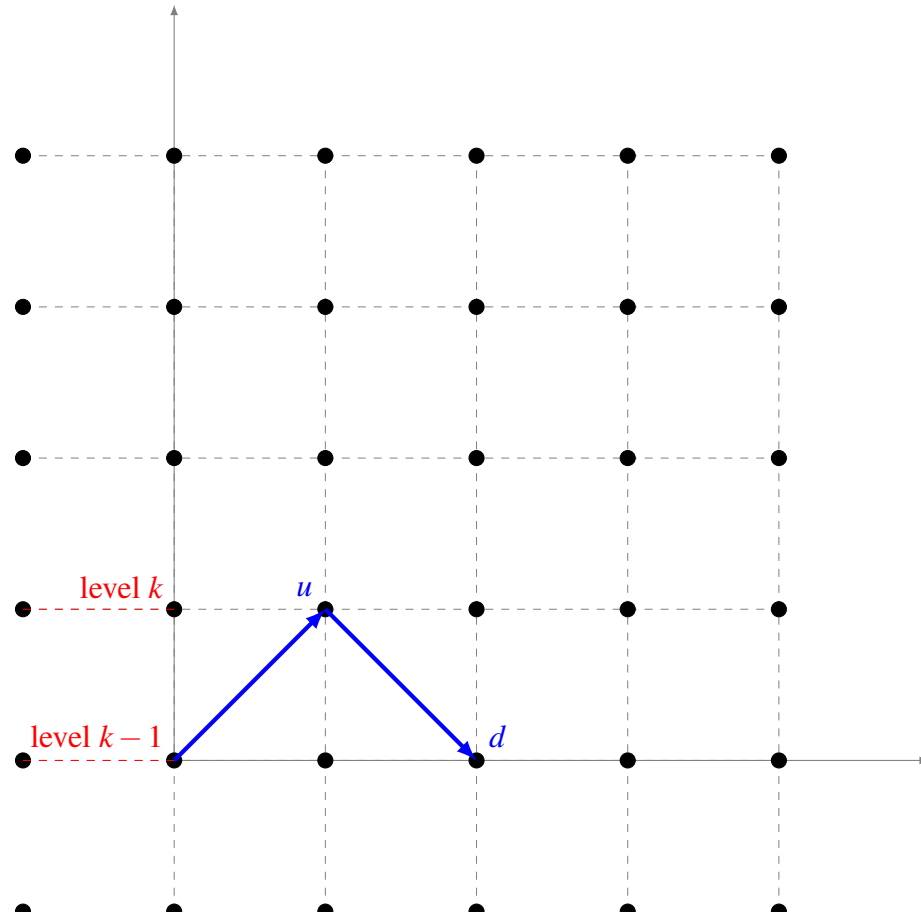
Figure 1: The *ud* peak at level *k* from Example 3.1

3.3   Dyck Paths: Generating Functions

We can enumerate the number of Dyck paths of semi-length n by finding its corresponding generating function, with the variable z encoding the parameter of semi-length, and performing coefficient extraction. In order to do so, we will develop its recurrence relation, solve it, and then deduce the coefficients into a well known sequence of numbers.

Every non-empty Dyck path can be decomposed as $\alpha = u\beta d\gamma$ or $\alpha = \beta u\gamma d$ where α, β, γ are Dyck paths, and β and γ may or may not be empty. This is called *first return decomposition*. (Deutsch, 1999)

First return decomposition allows us to decompose the set of n-Dyck paths as

$$\mathbf{D_n} = \mathbf{D_0}\hat{\mathbf{D}}_{\mathbf{n-1}} \cup \mathbf{D_1}\hat{\mathbf{D}}_{\mathbf{n-2}} \cup \cdots \cup \mathbf{D_{n-2}}\hat{\mathbf{D}}_{\mathbf{1}} \cup \mathbf{D_{n-1}}\hat{\mathbf{D}}_{\mathbf{0}} \tag{3.7}$$

and

$$\mathbf{D_n} = \hat{\mathbf{D}}_{\mathbf{0}}\mathbf{D_{n-1}} \cup \hat{\mathbf{D}}_{\mathbf{1}}\mathbf{D_{n-2}} \cup \cdots \cup \hat{\mathbf{D}}_{\mathbf{n-2}}\mathbf{D_1} \cup \hat{\mathbf{D}}_{\mathbf{n-1}}\mathbf{D_0} \tag{3.8}$$

for $n \geq 1$. Note that these are disjoint sets. Thus, (4.8) and (4.9) together imply that the cardinality of $\mathbf{D_n}$ is the sum,

$$|\mathbf{D_n}| = |\mathbf{D_0}||\mathbf{D_{n-1}}| + \cdots + |\mathbf{D_{n-1}}||\mathbf{D_0}| \tag{3.9}$$

where $n \geq 1$. (Deutsch, 1999)

Denote $D(z)$ as the generating function that counts the number of Dyck paths of semi-length n with z as the variable that codes the semi-length of the Dyck path. Let P_n denote the enumerating polynomial of each z^n. This is,

$$D(z) = \sum_{n=0}^{\infty} P_n z^n = \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n \tag{3.10}$$

since, clearly, the coefficient P_n in this case is $|\mathbf{D_n}|$. Similarly, $\hat{D}(z)$ is the generating function

$$\hat{D}(z) = \sum_{n=0}^{\infty} |\hat{\mathbf{D}}_{\mathbf{n}}| z^n \tag{3.11}$$

The parameter semi-length is what we refer to as additive. If we concatenate two Dyck paths together, the semi-length of the result is the sum of the semi-lengths of each Dyck path. Because of this, the enumerating polynomial for semi-length of the concatenation of two sets is the product between the enumerating polynomial of each set.

Consider (3.9). Taking the cardinality, multiplying both sides by z^n, and sum over each $n \geq 1$. This gives us,

$$\sum_{n=1}^{\infty} |\mathbf{D_n}| z^n = z \sum_{n=1}^{\infty} |\mathbf{D_0}||\mathbf{D_{n-1}}| z^{n-1} + z \sum_{n=1}^{\infty} |\mathbf{D_1}| z |\mathbf{D_{n-2}}| z^{n-2} +$$
$$\cdots + z \sum_{n=1}^{\infty} |\mathbf{D_{n-2}}| z^{n-2} |\mathbf{D_1}| z + z \sum_{n=1}^{\infty} |\mathbf{D_{n-1}}| z^{n-1} |\mathbf{D_0}| \tag{3.12}$$

where for each term where the index is less than zero is equal to 0. Note that $D(z) - 1 = \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n - 1 = \sum_{n=1}^{\infty} |\mathbf{D_n}| z^n$. Performing a shift of indices on the right hand side,

$$D(z) - 1 = z \sum_{n=0}^{\infty} |\mathbf{D_0}||\mathbf{D_n}| z^n + z \sum_{n=0}^{\infty} |\mathbf{D_1}||\mathbf{D_{n-1}}| z^n +$$
$$\cdots + z \sum_{n=0}^{\infty} |\mathbf{D_{n-1}}||\mathbf{D_1}| z^n + z \sum_{n=0}^{\infty} |\mathbf{D_n}||\mathbf{D_0}| z^n \tag{3.13}$$

Note that since for terms with coefficients with indices less than zero is equal to 0, then for $j > 0$, $\sum_{n=0}^{\infty} |\mathbf{D_{n-j}}| z^{n-j} = \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n$. For each n, this sum is

$$D(z) - 1 = z \Big(|\mathbf{D_0}| \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n + |\mathbf{D_1}| z \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n + |\mathbf{D_2}| z^2 \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n \cdots \Big)$$
$$D(z) - 1 = z \Big(\sum_{n=0}^{\infty} |\mathbf{D_n}| z^n \sum_{n=0}^{\infty} |\mathbf{D_n}| z^n \Big) = z D^2(z) \tag{3.14}$$

Therefore,

$$D(z) - 1 = z D^2(z) \tag{3.15}$$

An important feature of this derivation is the fact that we used (3.9) rather than (3.7) or (3.8) to derive it. If instead we used (3.7) then we would have that

$$D(z) - 1 = z D(z) \hat{D}(z) \tag{3.16}$$

9

Discussion: If $n = 0$, then there is one Dyck path, the empty path. That is, $|\mathbf{D_0}| = 1$. Otherwise, for any non-empty Dyck path there must be an initial rise and a first return step, i.e. first return decomposition. These up and down steps from the first return decomposition, as a pair, are a Dyck path of semi-length 1, which z enumerates. Between u and d is a Dyck path which may or may not be the empty path. That is, there is a Dyck path of elevation. Following the first return step is also a Dyck path, which may or may not be the empty path.
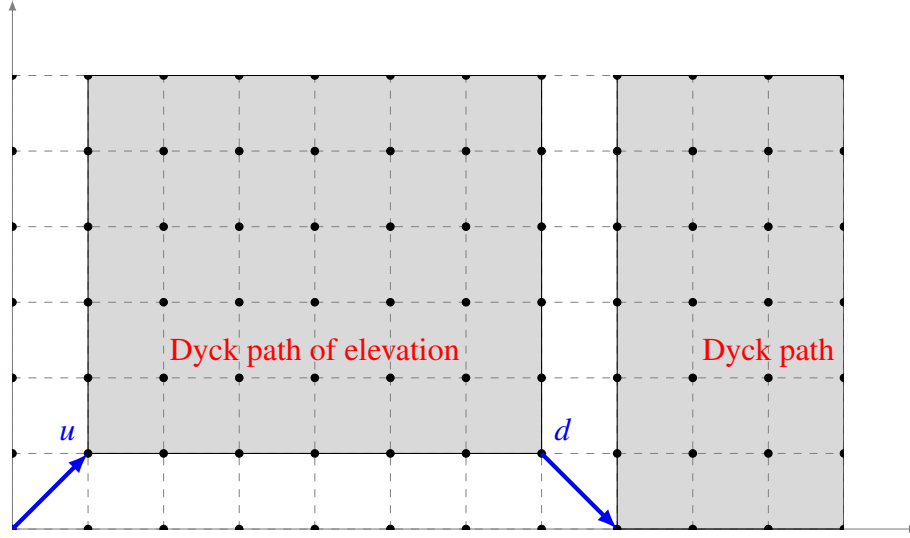

Figure 2: A visual of the recurrence relation we developed (4.16). For any non-empty Dyck path, there must be a rise from level 0 to 1 and a first return step at some point. In between that initial rise and first return step is a Dyck path where the 0 level is 1. After the first return step, there is also a Dyck path. Each of which may or may not be the empty path. If the Dyck path between the first rise and return step is empty, then they would form a peak at level 1. The horizontal separation is not literal and simply to aid visual representation.

Thus, the generating function for any non-empty ($n \geq 1$) Dyck path is $zD^2(z)$. That is,

$$D(z) - 1 = zD^2(z)$$

Now, adding the number of Dyck paths with length $n = 0$ to both sides (this is the empty case, 1), the recurrence relation that describes the total number of n-Dyck paths for $n \geq 0$ is,

$$D(z) = 1 + zD^2(z) \tag{3.17}$$

Rewriting this,

$$zD^2(z) - D(z) + 1 = 0 \tag{3.18}$$

This is a quadratic equation which we may easily solve for, obtaining,

$$D(z) = \frac{1 \pm \sqrt{1 - 4z}}{2z} \tag{3.19}$$

In order to to be sure which solution correctly enumerates Dyck paths, we take the Taylor series expansion of the discriminant. This gives us,

$$(1 - 4z)^{1/2} = 1 - 2z - 2z^2 - 4z^3 - \ldots \tag{3.20}$$

Since we are interested in the enumeration, we take the sign that will give us positive numbers. Therefore, we choose

$$D(z) = \frac{1 - \sqrt{1 - 4z}}{2z} \tag{3.21}$$

In order to utilize for the enumeration of the number of Dyck paths of semi-length n, we want to be able to represent $D(z)$ as a formal power series which we can perform coefficient extraction on. To do this, recall Netwon's generalized binomial theorem and perform the expansion on $(1 - 4z)^{1/2}$.

$$\sqrt{1 - 4z} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-4z)^n \tag{3.22}$$

The binomial coefficient, for any $n = 0, 1, 2, \ldots$, can be simplified as follows,

$$\binom{1/2}{n} = \frac{(\frac{1}{2})^n}{n!} = \frac{(\frac{1}{2})(\frac{1}{2} - 1)(\frac{1}{2} - 2) \cdots (\frac{1}{2} - n + 1)}{n!} = \frac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2}) \cdots (\frac{-2n+3}{2})}{n!}$$

$$= \frac{2^{-n}(-1)^{n-1}(1)(3) \cdots (2n - 3)}{n!} = (-1)^{n-1} \frac{(1)(3) \cdots (2n - 3)}{2^n n!} \tag{3.23}$$

Multiplying by $\frac{(2)(4) \cdots (2n-4)}{(2)(4) \cdots (2n-4)}$, we get

$$= (-1)^{n-1} \frac{(1)(2)(3)(4) \cdots (2n - 4)(2n - 3)}{2^n n!(2)(4) \cdots (2n - 4)} = (-1)^{n-1} \frac{(2n - 3)!}{2^n n! 2^{n-2}\big((1)(2) \cdots (n - 2)\big)}$$

$$= (-1)^{n-1} \frac{(2n - 3)!(2n - 2)(2n - 1)(2n)}{2^{2n-2} n!(n - 2)!(2n - 2)(2n - 1)(2n)} = (-1)^{n-1} \frac{(2n)!}{2^{2n} n!(n - 2)!(n - 1)n(2n - 1)}$$

$$= (-1)^{n-1} \frac{(2n)!}{4^n(2n - 1)n!n!} = \frac{(-1)^{n-1}}{4^n(2n - 1)} \binom{2n}{n} \tag{3.24}$$

That is, we have found that for $n = 0, 1, 2, \ldots$

$$\binom{1/2}{n} = \frac{(-1)^{n-1}}{4^n(2n - 1)} \binom{2n}{n} \tag{3.25}$$

Therefore, for $n = 0, 1, 2, \ldots$

$$\binom{1/2}{n}(-4z)^n = \frac{4^n(-1)^{2n-1}}{4^n(2n - 1)} \binom{2n}{n} z^n = \frac{-1}{2n - 1} \binom{2n}{n} z^n \tag{3.26}$$

Now, plugging this back into (3.22) the binomial expansion of $\sqrt{1 - 4z}$, we get,

$$\sqrt{1 - 4z} = \sum_{n=0}^{\infty} \frac{-1}{2n - 1} \binom{2n}{n} z^n = 1 - \sum_{n=1}^{\infty} \frac{1}{2n - 1} \binom{2n}{n} z^n \tag{3.27}$$

Now, substituting this expression into our expression $D(z)$, we get

$$D(z) = \frac{1 - \left(1 - \sum_{n=1}^{\infty} \frac{1}{2n-1} \binom{2n}{n} z^n\right)}{2z}$$

$$= \frac{\sum_{n=1}^{\infty} \frac{1}{2n-1}\binom{2n}{n} z^n}{2z} = \sum_{n=1}^{\infty} \frac{1}{2(2n-1)} \binom{2n}{n} z^{n-1} \tag{3.28}$$

substituting $n = n - 1$ and performing a shift of indices in the sum, we get

$$D(z) = \sum_{n=0} \frac{1}{2(2n+1)} \binom{2n+2}{n+1} z^n \tag{3.29}$$

Lastly,

$$\frac{1}{2(2n+1)} \binom{2n+2}{n+1} = \frac{1}{2(2n+1)} \frac{(2n+2)(2n+1)(2n)!}{(n+1)!(n+1)!}$$

$$= \frac{2n+2}{2(n+1)^2} \frac{(2n)!}{n!n!} = \frac{2(n+1)}{2(k+1)^2} \binom{2n}{n} = \frac{1}{n+1} \binom{2n}{n} \tag{3.30}$$

Recall that the nth Catalan number is $C_n = \frac{1}{n+1}\binom{2n}{n}$. That is, the coefficient of z^n is

$$\frac{1}{2(2n+1)} \binom{2n+2}{n+1} = \frac{1}{n+1} \binom{2n}{n} = C_n \tag{3.31}$$

Therefore, our generating function for enumerating the number of Dyck paths of semi-length n is

$$D(z) = \sum_{n=0}^{\infty} C_n z^n \tag{3.32}$$

The coefficient of each z^n is the nth Catalan number and is the number of distinct Dyck paths of semi-length n. That is, $[z^n]D(z) = C_n = |\mathbf{D_n}|$.

Result 3.1. The number of distinct Dyck paths of semi-length n is

$$C_n = \frac{1}{n+1} \binom{2n}{n} \tag{3.33}$$

That is, $|\mathbf{D_n}| = C_n$ where C_n is the nth Catalan number.

Result 3.2. The probability of a positive Dyck path occurring in the set of all Lattice paths from $(0,0)$ to $(0,2n)$, where $n \in \mathbb{N}$ and all paths are equiprobable, is,

$$\frac{1}{n+1} \tag{3.34}$$

Discussion: (A bijective approach) This enumeration can also be seen by noting that any Dyck word α must contain equally as many u's and d's while satisfying that at no time in the Dyck word there are more d's than u's. This is equivalent to the enumeration of Lattice paths to the point (n,n) below the diagonal line $y = x$. It is known that the number of such Lattice paths to the point (n,n) is C_n. The derivation comes from excluding all paths that don't satisfy such property from the total number of Lattice paths to (n,n). To find all paths that cross the $y = x$ line, it creates a bijection from them to a set of paths that can be enumerated. The details of this one proof can be seen in many other texts. It is important to understand it because the words that represent those such Lattice paths require the exact same property that we saw for Dyck words. However, generating functions allow enumeration beyond just the number of distinct paths of semi-length n as will be seen in the following section.

3.4   Bivariate Generating Functions: A Brief Walkthrough

In this section, we shall cover a specific example of enumerating more than one parameter of a Dyck path. Specifically, we will study semi-length as before in conjunction with the number of peaks. Following this section, we will conclude with a brief study of probabilities of various families of Dyck paths and Lattice paths under the assumption that all paths are equiprobable.

We let $p(\alpha)$ be the enumeration of a certain parameter of a Dyck path α. Denote P_n as the enumerating polynomial of a set of Dyck paths relative to a certain parameter p. This is,

$$P_n(t) = P_{\mathbf{D_n}}(t) = \sum_{\alpha \in \mathbf{D_n}} t^{p(\alpha)} \tag{3.35}$$

Similarly, $\hat{P}_n$ is defined as

$$\hat{P}_n(t) = P_{\hat{\mathbf{D}}_\mathbf{n}}(t) = \sum_{\hat{\alpha} \in \hat{\mathbf{D}}_\mathbf{n}} t^{p(\hat{\alpha})} \tag{3.36}$$

We will study two parameters. Semi-length as before and the number of peaks. z will code the semi-length and t will code the number of peaks.

The generating function $D(t,z)$ is a formal power series with $P_n(t)$ as the coefficient of each term z^n. This is,

$$D(t,z) = \sum_{n=0}^{\infty} P_n(t) z^n \tag{3.37}$$

The generating function $\hat{D}(t,z)$ is a formal power series with $\hat{P}_n(t)$ as the coefficient of each term z^n. This is,

$$\hat{D}(t,z) = \sum_{n=0}^{\infty} \hat{P}_n(t) z^n \tag{3.38}$$

In section 4.3 we developed the same generating function but instead with only semi-length as a parameter.

The parameter we will study is *additive*. By additive, we mean that if we concatenate two Dyck paths, then the value of the parameter of the result is simply the sum of the values of the parameters for each Dyck path we concatenated. That is, if we wish to study a parameter p, then for additive parameters,

$$p(\alpha\beta) = p(\alpha) + p(\beta)$$

where $p(\alpha)$ is the enumeration of the parameter p of the Dyck path α.

In the case of additive parameters, the number of distinct Dyck paths that satisfy the parameter is the product of the number of distinct paths for each set of Dyck paths. That is, if A and B are finite sets of Dyck paths satisfying parameter p, then

$$P_{AB}(t) = P_A(t) P_B(t)$$

(Deutsch, 1999)

Number of Peaks: The number of peaks in a Dyck path is additive and the number of peaks for any Dyck path of elevation of semi-length 0 is 1. That is, if $\hat{\alpha} \in \hat{\mathbf{D}}_\mathbf{0}$ then $p(\hat{\alpha}) = 1$.

Consider the enumerating polynomial for the number of peaks of Dyck paths of elevation. There are two cases for any Dyck path of elevation. Either $n = 0$ which gives us that the number

of peaks is 1. Or $n > 0$ and then the number of peaks of a Dyck path of elevation $\hat{\alpha} = u\alpha d$ is the number of peaks of the Dyck path α. That is,

$$\hat{P}_n(t) = \begin{cases} t & \text{if } n = 0 \\ P_n(t) & \text{if } n \geq 1 \end{cases} \tag{3.39}$$

For any Dyck path of elevation and semi-length $n = 0$, there is one peak, which t encodes. Otherwise, the number of peaks in all Dyck paths of elevation with semi-length $n \geq 1$ is the number peaks in all Dyck paths (not in elevation) with semi-length $n \geq 1$. This recursively defines $\hat{P}_n$ in terms of P_n. Now, multiplying by z^n, and summing over all n, we have

$$\sum_{n=0}^{\infty} \hat{P}_n(t)z^n = tz^0 + \sum_{n=1}^{\infty} P_n(t)z^n \tag{3.40}$$

$$\hat{D}(t,z) = t + D(t,z) - 1 \tag{3.41}$$

From (3.16) we know that $\hat{D}(z) = \frac{D(z)-1}{zD(z)}$. This implies that $\hat{D}(t,z) = \frac{D(t,z)-1}{zD(t,z)}$ since (4.14) is easily generalized as a bivariate generating function. Therefore,

$$\frac{D(t,z)-1}{zD(t,z)} = t + D(t,z) - 1 \tag{3.42}$$

multiplying both sides by $zD(t,z)$ we have

$$D(t,z) - 1 = tzD(t,z) + zD^2(t,z) - zD(t,z) \tag{3.43}$$

now subtracting $D(t,z)$ and adding 1 to both sides,

$$0 = zD^2(t,z) + tzD(t,z) - zD(t,z) - D(t,z) + 1$$

This is,

$$zD^2(t,z) - (1 + z - tz)D(t,z) + 1 = 0 \tag{3.44}$$

Clearly, this recurrence relation is in the form of a quadratic equation. Solving, we obtain,

$$\frac{1 + z - tz \pm \sqrt{(tz - z - 1)^2 - 4z}}{2z} \tag{3.45}$$

The Taylor series expansion of the discriminant is a series of negative numbers after the first term, similar to (4.20). Therefore, the solution is

$$\frac{1 + z - tz - \sqrt{(tz - z - 1)^2 - 4z}}{2z} \tag{3.46}$$

$$= \frac{1 + z - tz - \sqrt{1 - 2z + z^2 - 2tz - 2tz^2 + t^2z^2}}{2z} \tag{3.47}$$

The Narayana generating function is defined for $n \geq 1$ and its explicit expression $\rho(t,z)$ is defined as

$$\rho(t,z) = \frac{1 - z - tz - \sqrt{1 - 2z + z^2 - 2tz - 2tz^2 + t^2z^2}}{2tz} \tag{3.48}$$

14

Therefore, our solution to (4.40) is

$$D(t,z) = 1 + t\rho(t,z) \tag{3.49}$$

where for $n = 0$, $D(t,z) = 1$ and for $n \geq 1$, $D(t,z) = t\rho(t,z)$. The coefficients to the Narayana generating function are well known and are referred to as the Narayana numbers. They are denoted as

$$v_{n,k} = [t^k z^n]\rho(t,z) \tag{3.50}$$

We reference Appendix D.5 from *Dyck path enumeration* by Deutsch (Deutsch, 1999) to define the Narayana numbers as,

$$v_{n,k} = \frac{1}{n}\binom{n}{k}\binom{n}{k+1} \tag{3.51}$$

for $n \geq 1$. Note that $v_{n,k}$ is positive only for $1 \leq k \leq n$ and otherwise zero. Therefore,

$$[t^k z^n]D(t,z) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1} \tag{3.52}$$

for $n \geq 1$. If $n = 0$, then $k = 0$, and we define $[t^k z^n]D(t,z) = 1$.

3.5 Equiprobable Paths

Now, we will utilize our enumeration of Dyck paths and Lattice paths to study the probabilities of various paths given certain sample spaces and/or conditions. We will assume unless explicitly stated otherwise that all paths are equiprobable in whatever sample space we are studying.

3.5.1 Example

Suppose that two government candidates are in a fair election race. Let B be the other party. Assume that the probability of A gaining a vote is $p = \frac{1}{2}$. The results end in a tie. Given that $10,000$ ballots came in...

[leftmargin=20pt]What is the probability that one of the candidates did not allow the other to ever get a lead during the entire election race?Let Ω denote all possible election results that result in a tie and let Z denote the event of this particular result happening. Let A be one candidate and B be the other. For this problem, either candidate could have been the one to hold lead during the election race. Each case is disjoint and equiprobable. Denote Z_A as the event that A never allowed B to have the lead and Z_B as the event that B never allowed A to have the lead. The events are equal in cardinality. Then $Z = Z_A \cup Z_B$ and since these events are clearly disjoint and equal in cardinality,

$$\mathbb{P}(Z) = \mathbb{P}(Z_A) + \mathbb{P}(Z_B) = \frac{|Z_A|}{|\Omega|} + \frac{|Z_B|}{|\Omega|} = 2\frac{|Z_A|}{|\Omega|}$$

Let S_m be a simple random walk where $m =$ number of ballots counted $= 10,000$. If m were odd then clearly one of the candidates must have more votes than the other and the probability would be $\mathbb{P}(Z) = 0$. Let n denote the semi-length of the random walk. In other words, $2n = m$. Initially, the election starts fair. That is, $S_0 = 0$. Each ballot casted will represent a discrete step in time. If a ballot is for A, then we let it add 1 to the current state of the system with probability $p = \frac{1}{2}$. Otherwise, we subtract 1 with probability $1 - p = \frac{1}{2}$.

$|\Omega|$ is the total number of simple random walks that terminate at 0. We know then that from the proof of Theorem 2.1 and the total number of diagonal Lattice paths (3.4),

$$|\Omega| = \binom{2n}{n}$$

$|Z_A|$ is the total number of simple random walks that terminate at 0 where $S_i \geq 0$ for all $i = 0, 1, 2, \ldots, m$. In other words, it is a simple random walk that ends at 0 where at no point in time i are there more down steps (ballots for B) than there are up steps. This is a Dyck path of semi-length n. Thus,

$$|Z_A| = C_n = \frac{1}{n+1} \binom{2n}{n}$$

and we have that,

$$\mathbb{P}(Z) = 2\frac{|Z_A|}{|\Omega|} = 2\frac{C_n}{\binom{2n}{n}} = 2\frac{1}{n+1} = \frac{2}{5001}$$

$$\mathbb{P}(Z) \approx 0.03999\%$$

(b) What is the probability that there were exactly $2,000$ number of times that the ballots shifted from A to B, given that A never allowed B to take the lead at any point? By shifted from A to B, we mean that the last ballot casted was for A and the next was casted for B. What about only 1 shift from A to B?

Solution 2. Let V denote the event that A never allowed B to take the lead but resulted in a tie and had $2,000$ shifts in the election. Each shift represents a peak in the election that is a Dyck path as discussed in part (a). Additionally, we already found $|Z_A|$ Therefore,

$$\mathbb{P}(V) = \frac{|V|}{|Z_A|} = \frac{v_{5000,2000}}{C_{5,000}}$$

Unfortunately, this isn't easily solvable by just plugging in numbers and evaluating. It is too large of a number in both the numerator and denominator for much practical use itself, even though the numbers are well known and have been calculated. In fact, the 5000th Catalan number is well over the number of observable particles in the universe. Likewise for $v_{5000,2000}$.

For the second part of this question, this can be easily solved by noting that if there is one peak in a Dyck path of semi-length 5000 then that peak must occur at level 5000 and thus, since there is only one such path, the probability in that case is

$$\frac{1}{C_{5000}}$$

Acknowledgments

References

[1] Deutsch, E. (1999). Dyck Path enumeration. Discrete Mathematics, 204(1-3), 167–202.

[2] Billingsley, P. (1995). Section 7. In Probability and measure (pp. 92–101). essay, Wiley.

[3] Grimmett, G., & A., W. D. J. (2017). Chapter 10. In Probability an introduction (2nd ed., pp. 167–174). essay, Oxford University Press.

[4] Hoel, P. G., Port, S. C., & Stone, C. J. (1987). Chapter 1. In Introduction to stochastic processes (pp. 1–7). essay, Waveland Press.