

1. **The Random Walk on the Integers:** We start with the **coin toss game** which uses a coin with probability \mathcal{P} for Head and probability $q = 1 - \mathcal{P}$ for Tail. You get a dollar if it comes up Head and lose a dollar if it shows Tail. The coin will be tossed infinitely many times and the random variable X_i measures your earnings from the i^{th} toss:

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ toss is Head} \\ -1 & \text{otherwise} \end{cases}$$

Your profit (excess of heads over tails) after m tosses is

$$S_m = X_1 + \cdots + X_m.$$

Your progress in this game may be thought of as describing **a random walk on the integers**. One way to think of this is *on the line*, or in one dimension: Let $S_0 = 0$ signify the start at the origin. Each toss is a *step* in which “Head” moves you one integer to the right from the current position and “Tail” moves one integer to the left. Thus S_j is the position of the walk after the j^{th} step (toss). The study of this interesting game reveals some surprising properties, and it is another very significant example showing the power of generating functions.

- **Positive Paths - The Ballot Theorem** Another way to depict the above walk is in two dimensions, i.e., in the plane: we use the x-axis for “time” and the y-axis for the progress of the walk. The sequence of points $\{(0, 0), (1, S_1), (2, S_2) \dots, (m, S_m)\}$ describes the start at $(0, 0)$ [at time zero the game is even]; after step j our point has x -coordinate j and y -coordinate $S_j = X_1 + \cdots + X_j$, the *excess* of heads over tails for the first j tosses. If we connect successive points by straight line segments we get what we will call a **random walk path** with m steps (starting at the origin, ending at (m, S_m)). There are 2^m such paths (because each sequence of m tosses gives a distinct path). For now we will take $\mathcal{P} = 1/2$, so each of the 2^m paths is equally likely.

We write

$$m = n_H + n_T$$

for the number of tosses, n_H denoting the number of Heads and n_T the number of Tails, and observe that the score (our fortune) after m tosses is the difference between the number of Heads and the number of Tails, or

$$S_m = n_H - n_T.$$

We fix a positive integer k , $0 < k \leq m$, and focus our attention on the paths that go from $(0, 0)$ to (m, k) ; i.e. $S_m = k$ is the score after m steps. Since $S_m = n_H - n_T = 2n_H - m$, k is even if m is, and vice-versa. Also note that the number of distinct paths which start at $(0, 0)$ and end at (m, k) is

$$N_{m,k} = \binom{m}{n_H},$$

each one equally likely. A path is called **GOOD** (positive) if $S_j > 0$, for all $j, 1 \leq j \leq m$; i.e., it remains positive after the start. The ballot theorem says that the number of good paths from $(0, 0)$ to (m, k) is

$$G_{m,k} = \left(\frac{n_H - n_T}{m} \right) N_{m,k} = \frac{k}{m} \binom{m}{n_H} \quad (1)$$

This gives k/m as the probability that a path from $(0, 0)$ to (m, k) is good. The statement in (1) is called the Ballot Theorem because if Candidate A gets n_H votes and Candidate B gets $n_T < n_H$ votes, (there are $m = n_H + n_T$ ballots and A wins by $k = n_H - n_T$ votes) the probability k/m describes the chances that A leads throughout the counting of the ballots, where we assume that the m ballots are counted one at a time, and each of the $m!$ ordering of the ballots is equally likely. It is interesting to note that this probability is the slope of the line joining $(0, 0)$ to (m, k) . Note also that if candidate A wins by only a small margin, the probability that he was always ahead during the counting is small¹.

- **Non-negative Paths:** We will apply the Ballot Theorem to the paths that start at $(-1, -1)$ and end at (m, k) , where now $k \geq 0$ (i.e., zero is allowed). We call a path “pseudo-positive” if, after the first toss, it is strictly above $y = -1$. Every such path must pass through $(0, 0)$, and thereafter, it is a true nonnegative path from $(0, 0)$ to (m, k) ; i.e., it’s on or above the x-axis. These so-called “positive” paths from $(-1, -1)$ have $m + 1$ steps, have $n_H + 1$ Heads, and have n_T Tails, so (1) gives

$$NN_{m,k} = \frac{k + 1}{m + 1} \binom{m + 1}{n_H + 1} \quad (2)$$

as the number of **non-negative paths** from $(0, 0)$ to (m, k) . Dividing by $N_{m,k}$ reveals that the probability of a non-negative path from $(0, 0)$ to (m, k) is

$$\frac{k + 1}{n_H + 1}.$$

An interesting special case of (2) is when $n_H = n_T = n$, so $k = 0$. and $m = 2n$. This gives

$$NN_{2n,0} = \frac{1}{2n + 1} \binom{2n + 1}{n + 1} = \frac{1}{n + 1} \binom{2n}{n},$$

the last expression easily deduced from the middle one. This relation reveals that the set of n node binary trees and the set of non-negative random walk paths from $(0, 0)$ to $(2n, 0)$ have **the same size!!**

- **First Lead of \$1:** Write \mathcal{P} for the probability of Head, no longer assumed to be $1/2$. We study the probability of the event A_n that the walk *first* arrives at the integer 1 at the n^{th} step (you lead by 1 after step n but were even, or behind, previously). Thus

$$A_n = \{S_n = 1, \text{ and } S_j < 1, j < n\}.$$

¹In the 2000 election in Florida, Bush got 535 more votes than Gore, out of more than 5.8 million votes cast, a tiny margin of victory. In fact he was many times reported to be trailing in the count.

Let X be a random variable (on the sample space of coin tosses) that counts the number of trials you need to first achieve a lead of 1 over the opponent (who gets \$1 for Tail and loses \$1 for Head). A_n is the event $\{X = n\}$ and we write

$$p_n = f_X(n) = \text{Prob}(X = n)$$

for its probability. Clearly $p_1 = \mathcal{P}$ (Head on first toss) and $p_2 = 0$ (HH, HT, TH, TT gives $S_2 = 2, 0, 0, -2$, respectively). In fact $p_{2j} = 0$ since it is impossible to lead by 1 after an even number of trials). Also $p_3 = (1 - \mathcal{P})\mathcal{P}^2$ and $p_5 = 2(1 - \mathcal{P})^2\mathcal{P}^3$ (THTHH or TTHHH). We will compute the generating function of X ,

$$\phi_X(s) = \sum_{k=1}^{\infty} p_k s^k.$$

We already observed that $p_k = 0$ if k is even. Take $n > 1$ and note that $\{X = n\}$ occurs only, if for some $k, 2 \leq k < n$,

- (a) the first trial fails (this is phase I)
- (b) you first equalize Heads and Tails at trial k (phase II comprises trials $2, \dots, k$)
- (c) you take $n - k$ additional trials to first get a lead of 1 (phase III covers trials $k + 1, \dots, n$)

Clearly (a) is necessary for A_n , and if you are trailing by 1 at time 1 and will be leading by 1 at time n , you must be equal sometime; k is when that first occurs.

Let X_1 measure the number of trials (after trial 1) for phase II and X_2 the number of trials for phase III. Letting C be “tail on trial 1”, we have

$$\{X = n\} = \bigcup_{k=2}^{n-1} (C \cap \{X_1 = k - 1\} \cap \{X_2 = n - k\})$$

and because the events in the union are mutually exclusive,

$$p_n = \sum_{k=2}^{n-1} \text{Prob}(C \cap \{X_1 = k - 1\} \cap \{X_2 = n - k\}).$$

The three events in the intersections are independent because the phases are non-overlapping Bernoulli trials, and since the X_i 's, like X , measure the number of trials needed to first get a lead of 1,

$$p_n = q \sum_{k=2}^{n-1} p_{k-1} p_{n-k} = q(p_1 p_{n-2} + \dots + p_{n-2} p_1)$$

Let $\{a_i\} = \{p_i\} * \{p_i\}$ be the convolution of $\{p_i\}$ with $\{p_i\}$ (so $A(s) = (\phi_X(s))^2$) and note that the expression in parentheses above, is a_{n-1} , so

$$p_n = q a_{n-1}.$$

Multiply this equation by s^n and sum the terms from $n = 2$ to ∞ to see

$$\sum_{n=2}^{\infty} p_n s^n = qs \sum_{n=2}^{\infty} a_{n-1} s^{n-1}.$$

The expression on the left is $\phi_X(s)$, except the first term for $n = 1$ is missing, namely $p_1 s$. Since $p_1 = \mathcal{P}$,

$$\phi_X(s) - \mathcal{P}s = qsA(s) = qs(\phi_X(s))^2,$$

the last equality because $\{a_i\}$ is the convolution of $\{p_i\}$ with itself. This is a quadratic equation (in ϕ_X) with solutions

$$\phi_X(s) = \frac{1 \pm \sqrt{1 - 4\mathcal{P}qs^2}}{2qs}$$

and we can eliminate the “+” solution because it has infinite limit as $s \rightarrow 0$. Evaluate this expression at $s = 1$ and use $q = 1 - \mathcal{P}$, to see that

$$\phi_X(1) = \frac{1 - |1 - 2\mathcal{P}|}{2q}.$$

If $\mathcal{P} < 1/2$, $|1 - 2\mathcal{P}| = (1 - 2\mathcal{P})$ and $\phi_X(1) = \mathcal{P}/q < 1$. If $\mathcal{P} \geq 1/2$, $|1 - 2\mathcal{P}| = 2\mathcal{P} - 1$ and $\phi_X(1) = 1$. Therefore

$$\phi_X(1) = \sum_{n=1}^{\infty} p_n = \begin{cases} \frac{\mathcal{P}}{q} & \text{if } \mathcal{P} < \frac{1}{2} \\ 1 & \text{if } \mathcal{P} \geq \frac{1}{2} \end{cases} \quad (3)$$

$\phi_X(1)$ is the sum of the probabilities that we first lead by 1 after *one* step, after *three* steps, after *five* steps, etc.; i.e., the probability that we at some time get a lead of one over the opponent. According to (3) if the game is biased against us (i.e., if $\mathcal{P} < 1/2$), this probability is $\mathcal{P}/q < 1$ and there is a positive chance - $1 - \mathcal{P}/q$ - that we never get one unit ahead **SURPRISE 1!** On the other hand we are sure to eventually get a lead of 1 if the coin is fair or biased in our favor.

Now we look at $E(X)$. If $\mathcal{P} < 1/2$, $E(X) = \infty$ since, by the previous fact, $X = \infty$ with positive probability. So we may assume $\mathcal{P} \geq 1/2$. Differentiating $\phi_X(s) = (1 - \sqrt{1 - 4\mathcal{P}qs^2})/(2qs)$ with respect to s and evaluating the result at $s = 1$ gives

$$E(X) = \phi'_X(s)|_{s=1} = \frac{1}{2\mathcal{P} - 1}.$$

This is infinite for the fair coin with $\mathcal{P} = 1/2$ and gives **SURPRISE 2:** though we are certain to get a lead of 1, we expect to wait infinitely long for it to happen!.

Now lets assume the coin is fair so $\mathcal{P} = 1/2$: We note that the previous observations show that *the walk is recurrent*: We are certain to gain 1. After this occurs, we are certain to gain 1 again, etc., so we eventually get to each positive integer. By symmetry, we are also certain to get to every negative integer. In fact these remarks really say that if the coin is fair, **we visit every integer infinitely many times**. On the other hand, and somewhat paradoxically, we expect an infinite wait to arrive at any *particular* integer.

- **Discussion:** We obtained the generating function

$$\phi_X(s) = \sum_{k=1}^{\infty} p_k s^k = \frac{1 - \sqrt{1 - 4\mathcal{P}qs^2}}{2qs} \quad (4)$$

by a structural argument, and by using convolution. In fact we do know that $p_k = 0$ if k is even, since you can't lead by 1 after an even number of tosses. Also we can easily show that

$$p_{2j+1} = \left[\frac{1}{2j+1} \binom{2j}{j} \right] \mathcal{P}^{j+1} q^j \quad (5)$$

for $j = 0, 1, \dots, \infty$. The expression in square brackets is the number of non-negative paths from $(0, 0)$ to $(2j, 0)$, and by symmetry this must also be the number of non-positive paths. Any such non-positive path must have j Heads and j Tails, and then end with a final Head, so it can arrive at $(2j + 1, 1)$. Thus *every* path that first becomes positive at step $2j + 1$ has (product) probability $\mathcal{P}^{j+1}(1 - \mathcal{P})^j$, and there are $\binom{2j}{j}/(2j + 1)$ of them. However it is not clear how, at this point, to use (5) in the middle sum in (4) so as to obtain the convenient closed form for $\phi_X(s)$ that is shown in the final expression in (4). It was this closed form and basic properties of generating functions that allowed us to learn the interesting and surprising facts about the behaviour of the random walk on the integers.