Course:	Introduction to Stochastic Processes
Term:	Fall 2019
Instructor:	Gordan Žitković

Lecture 3
Random Walks

3.1 Stochastic Processes

Definition 3.1.1. A **stochastic process** is a sequence - finite or infinite - of random variables.

We usually write $\{X_n\}_{n\in\mathbb{N}_0}$ of $\{X_n\}_{0\leq n\leq T}$, depending on whether we are talking about an infinite or a finite sequence. The number $T \in \mathbb{N}_0$ is called the **(time) horizon**, and we sometimes set $T = +\infty$ when the sequence is infinite. The index n is often interpreted as *time*, so that a stochastic process can be thought of as a model of a random process evolving in time. The initial value of the index n is often normalized to 0, even though other values - such as 1 - are also used. This it usually very clear from the context.

It is important that all the random variables $X_0, X_1, \dots$ "live" on the same sample space. This way, we can talk about the notion of a **trajectory** or **sample path** of a stochastic process: it is, simply, the sequence of numbers

$$X_0(\omega), X_1(\omega), \dots$$

but with ω considered "fixed". In other words, we can think of a stochastic process as a random variable whose values are not numbers, but sequences of numbers. This will become much clearer once we introduce enough examples.

3.2 The Simple Symmetric Random Walk

Definition 3.2.1. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ is said to be a **simple symmetric random walk** if

1. $X_0 = 0$,
2. the random variables $\delta_1 = X_1 - X_0$, $\delta_2 = X_2 - X_1$, ... are independent

3. each δ_n is a **coin toss**, i.e., its distribution is given by

$$\delta_n \sim \begin{array}{c|cc} & -1 & 1 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Remark 3.2.2.

1. Definition 3.2.1 captures the main features of an idealized notion of a particle that gets shoved, randomly, in one of two possible directions. These "shoves" are modeled by the random variables $\delta_1, \delta_2, \dots$ and the position of the particle after n "shoves" is X_n; indeed,

$$X_n = \delta_1 + \delta_2 + \cdots + \delta_n, \text{ for } n \in \mathbb{N}.$$

   It is important to assume that any two "shoves" are independent of each other; the most important properties of random walks depend on this in a critical way.

2. Sometimes, we only need a finite number of steps of a random walk, so we only care about the random variables $X_1, \dots, X_T$. This stochastic process (now with a finite time horizon T) will also be called a random walk, and it should be clear from the context whether we need a finite or infinite horizon.

3. The starting point $X_0 = 0$ is just a normalization. Sometimes we need more flexibility and allow our process to start at $X_0 = x$ for some $x \in \mathbb{N}$. To stress that fact, we talk about the random walk **started at** x. If no starting point is mentioned, you should assume $X_0 = 0$.

4. We will talk about the **biased** or **assymetric** random walks a bit later. The only difference will be that the probabilities of each δ_n taking values 1 or -1 will not be $\frac{1}{2}$ (but will also not change from step to step).

We defined a notion of a sample path (or a trajectory) of a stochastic process. For a random walk on a finite horizon T, a trajectory is simply a sequence of natural numbers starting from 0. Different realizations of the coin-tosses δ_n will lead to different trajectories, but not every sequence of natural numbers corresponds to a trajectory. For example $(0, 3, 4, 5)$ is not a sample path of a random walk because X_1 can only take values 1 or -1. In fact, a finite sequence $(x_0, x_1, \dots, x_T)$ is a (possible) sample path of a random walk if and only if $x_0 = 0$ and $x_k - x_{k-1} \in \{-1, 1\}$ for each k.

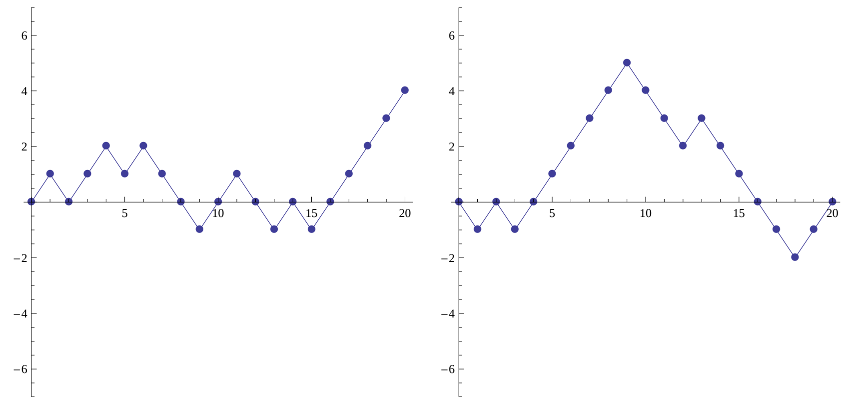

Figure 1. Two sample paths of a random walk

Two figures above show two different trajectories of a simple random walk. Each one corresponds to a (different) frozen $\omega \in \Omega$, with n going from 0 to 20.

Unlike in Figure 1. above, Figure 2. below shows two "time slices" of the same random process; in each graph, the time t is fixed ($n = 15$ vs. $n = 25$) but the various values random variables X_{15} and X_{25} can take are presented through the probability mass functions.

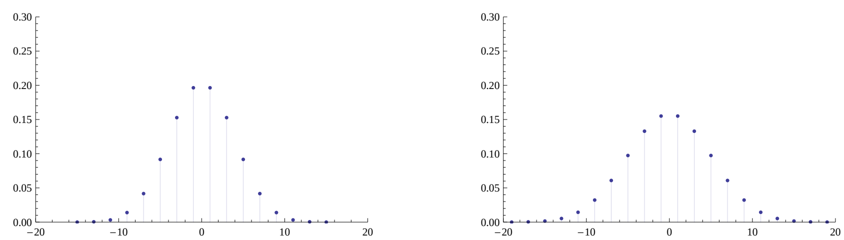

Figure 2. probability mass functions of X_n, corresponding to $n = 15$ and $n = 20$.

3.3 The canonical probability space

Let us build a sample space on which a random walk with a finite horizon T can be constructed. Since the basic building blocks of the random walk are its increments, it makes sense to define the elementary outcome ω as a sequence, of size T, consisting of 1s and -1s. More precisely, we take

$$\begin{aligned}\Omega &= \{(c_1, \dots, c_T) : c_1, \dots, c_T \in \{-1, 1\}\} \\ &= \underbrace{\{-1, 1\} \times \cdots \times \{-1, 1\}}_{T \text{ times}} = \{-1, 1\}^T.\end{aligned}$$

What probability would be appropriate here? That is easy, too, since the individual coin tosses need to be fair and independent. This dictates that any two sequences be equally likely. There are 2^T possible sequences of length T of 1s and -1, so $\mathbb{P}[\{\omega\}] = 2^{-T}$, for each ω.

This way, given an $\omega = (c_1, \dots, c_n) \in \Omega$ and $n = 1, \dots, T$, we define

$$\delta_n(\omega) = c_n \text{ and } X_n(\omega) = c_1 + \cdots + c_n,$$

with $X_0(\omega) = 0$. It is easy to see that $\{X_n\}$ is, indeed, a random walk.

This setup also makes the computation of various probabilities that have to do with the random walk easy (at least in theory) - we simply count the number of sequences of coin tosses that correspond to the event in question, and then divide by 2^T.

Example 3.3.1. Let us compute the probability that $X_n = k$, for some $n \in 0, \dots, T$ and $k \in \mathbb{N}$. For this, we first need to count the number of $\omega \in \Omega$ for which $X_n(\omega) = k$. Using the definition above, this is equivalent to counting the number of n-tuples $(c_1, \dots, c_T)$ of 1s and -1s such that $c_1 + c_2 + \cdots + c_n = k$.

To solve that problem, we note that, in order for n numbers, each of which is either 1 or -1, to add up to k, we must have $n_1 - n_{-1} = k$, where n_1 is the number of 1s and n_{-1} the number of -1s in the sequence $(c_1, \dots, c_n)$. On the other hand, it is always the case that $n_1 + n_{-1} = n$, so we must have

$$n_1 = \tfrac{1}{2}(n + k) \text{ and } n_{-1} = \tfrac{1}{2}(n - k).$$

The first thing we observe that this is not possible unless $n + k$ (and then also $n - k$) is an even number. If you think about it for a second, it makes perfect sense. The random walk can only take odd values at odd times and even values at even times. Hence, $\mathbb{P}[X_n = k] = 0$ if n and k have different parity (one is odd and the other even).

Suppose now that $n + k$ is even. Then $n - k$ is also even, and so n_1 and n_{-1}, defined above, are natural numbers. Our question then becomes

> How many sequences $(c_1, c_2, \dots, c_T)$ with values in $\{-1, 1\}$ are there such that there are exactly n_1 1s among the first n elements?

We can pick the positions of n_1 1s among the first n spots in $\binom{n}{n_1}$ ways, and fill the remaining n_{-1} slots by -1s. The values at positions between $n+1$ and T do not affect the value of X_n, so they can be chosen arbitrarily, in 2^{T-n} ways. Therefore, the answer to our question is $\binom{n}{n_1}2^{T-n}$.

To turn the count into a probability, we need to divide by 2^T. All in all, we get the following result: for $n \in \{0, \dots, T\}$ and $-n \le k \le n$, we have

$$\mathbb{P}[X_n = k] = \begin{cases} \binom{n}{(n+k)/2}2^{-n}, & \text{if } n+k \text{ is even, and} \\ 0, & \text{otherwise.} \end{cases}$$

Another approach to this questions would be via the fact that

$$\tfrac{1}{2}(X_n + n) = \tfrac{1}{2}(\delta_1 + 1) + \tfrac{1}{2}(\delta_2 + 1) + \dots + \tfrac{1}{2}(\delta_n + 1).$$

Since δ_m takes values ± 1 with probability $\frac{1}{2}$, each, $Y_m = \frac{1}{2}(\delta_m + 1)$ is a Bernoulli random variable, with parameter $p = \frac{1}{2}$. They are independent (since δs are), and, so, $\frac{1}{2}(X_n + n)$ has the binomial distribution, with parameters n and $\frac{1}{2}$. It follows that for $-n \le k \le n$ and n and k of the same parity, we have

$$\mathbb{P}[X_n = k] = \mathbb{P}[\tfrac{1}{2}(X_n + n) = \tfrac{1}{2}(n+k)] = \binom{n}{(n+k)/2}2^{-n}.$$

3.4 Biased random walks

If the steps of the random walk preferred one direction to the other, the definition would need to be tweaked a little bit:

Definition 3.4.1. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ is said to be a **(simple) biased random walk** with parameter $p \in (0,1)$ if

1. $X_0 = 0$,
2. the random variables $\delta_1 = X_1 - X_0$, $\delta_2 = X_2 - X_1$, ... are independent
3. each δ_n is a p**-biased coin toss**, i.e., its distribution is given by

$$\delta_n \sim \begin{array}{c|cc} & -1 & 1 \\ \hline & 1-p & p \end{array}$$

Could we reuse the sample space Ω to build a biased random walk? Yes, we could, but we would need to assign different probabilities to elementary outcomes. Indeed, if $p = 0.99$, the probability that all the increments δ take the value $+1$ is larger than the probability that all steps take the value -1.

More generally, the sequence $\omega = (c_1, \dots, c_N)$ consisting of n_1 1s and n_{-1} -1s should be assigned the probability $p^{n_1}(1-p)^{n_{-1}}$. Only then will δs be independent and distributed as $\begin{array}{c|cc} & -1 & 1 \\ \hline & 1-p & p \end{array}$, as required.

We can still use this sample space to figure out distributions of various random variables, but we cannot always simply "count and divide by the size of Ω" like we could when $p = \frac{1}{2}$. Sometimes it still works, as the following example shows:

Example 3.4.2. Let us try to compute the same probability as in Example 3.3.1 above, namely $\mathbb{P}[X_n = k]$, but now in the biased case.

To simplify our lives, we can assume without loss of generality that $n = N$, i.e., that nothing happens after n. We still need to identify those $\omega = (c_1, \dots, c_T)$ for which $X_n(\omega) = k$, and it turns out that the reasoning is the same as in the symmetric case. We need exactly $n_1 = \frac{1}{2}(n+k)$ of the cs to be equal to 1 and exactly $n_{-1} = n - n_1$ of them to be equal to -1. The lucky break is that each sequence with exactly n_1 1s carries the same probability, namely $p^{n_1}(1-p)^{n_{-1}}$, no matter where these 1s are. In other words, it just happened that all ω in the event $\{X_n = k\}$ have the same probability. Therefore, the probability of $\{X_n = k\}$ is simply $p^{n_1}(1-p)^{n-n_1}$ multiplied by the number of ωs that constitute it. We have already computed that in Example 3.3.1 - the answer is $\binom{n}{n_1}$ - and, so

$$\mathbb{P}[X_n = k] = \begin{cases} \binom{n}{(n+k)/2} \, p^{(n+k)/2}(1-p)^{(n-k)/2}, & \text{if } n+k \text{ is even, and} \\ 0, & \text{otherwise.} \end{cases}$$

As in the symmetric case, this also follows from the fact that $\frac{1}{2}(X_n + n)$ is binomial, with parameters n and p.

3.5 The Reflection Principle

Counting trajectories in order to compute probabilities can be quite powerful, as the following example shows. It also reveals a potential weakness of the combinatorial approach: it works best when all ω are equally likely (i.e., when $p = \frac{1}{2}$ in the case of the random walk).

We start by asking a simple question; what is the typical record value of the random walk, i.e., how far "up" does it typically get? Clearly, the largest value it can attain is T at time T, provided that all coin tosses came up $+1$. This is, however, extremely unlikely - it happens with probability 2^{-T}. On the other hand, this maximal value is at last 0, since $X_0 = 0$, already. A bit of thought reveals that any value between those two extremes is possible, but it is not at all easy to compute their probabilities.

More precisely, if $\{X_n\}$ is a simple random walk with time horizon T. We define the **running maximum process** $\{M_n\}_{n\in\mathbb{N}_0}$ by

$$M_n = \max(X_0, \dots, X_n), \quad \text{for } 0 \le n \le T.$$

It turns out that a nice counting trick - known as the **reflection principle** - can help us compute the distribution of M_n for each n.

Proposition 3.5.1. *Let $\{X_n\}_{0\le n\le T}$ be a simple symmetric random walk. For $1 \le n \le T$, the support of the random variable $M_n = \max(X_0, \dots, X_n)$ is $\{0, 1, \dots, n\}$ and its probability mass function is given by*

$$\begin{aligned}\mathbb{P}[M_n = k] &= \mathbb{P}[X_n = k] + \mathbb{P}[X_n = k+1] \\ &= \binom{n}{\lfloor \frac{n+k+1}{2} \rfloor} 2^{-n}, \text{ for } k = 0, \dots, n,\end{aligned}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x.

Proof. As usual, we may assume without loss of generality that $n = T$ since the values of $\delta_{n+1}, \dots, \delta_T$ do not affect M_n at all.

We start by picking a level $l \in \{0, 1, \dots, n\}$ and first compute the probability $\mathbb{P}[M_n \ge l]$. The symmetry assumption ensures that all trajectories are equally likely, so we can do this by counting the number of trajectories whose maximal level reached is at least l, and then multiply by 2^{-n}.

What makes the computation of $\mathbb{P}[M_n \ge l]$ a bit easier than that of $\mathbb{P}[M_n = l]$ is the following equivalence

$$M_n \ge l \quad \text{if and only if} \quad X_k = l \text{ for some } k.$$

In words, the set of trajectories whose maximum is at least l is exactly the same as the set of trajectories that hit the level l at some time. Let us denote the set of ω with this property by A_l, so that $\mathbb{P}[M_n \ge l] = \mathbb{P}[A_l]$.

We can further split A_l into three disjoint events $A_l^>$, $A_l^=$ and $A_l^<$, depending on whether $X_n < l$, $X_n = l$ and $X_n > l$. The idea behind the reflection principle is that $A_l^>$ and $A_l^<$ have exactly the same number of elements. To see that that is, indeed, true, we is take an $\omega \in A_l^>$ and denote by $F(\omega)$ the *first time* the corresponding trajectory visits the level l. After that, we flip the portion the trajectory between $F(\omega) + 1$ and n around the level l. In terms of ω, this amounts to flipping the signs of its last $n - F(\omega) + 1$ entries (see Figure 3 below). It is easy to see that this establishes a bijection between the sets $A_l^>$ and $A_l^<$, making these two sets equal in size.

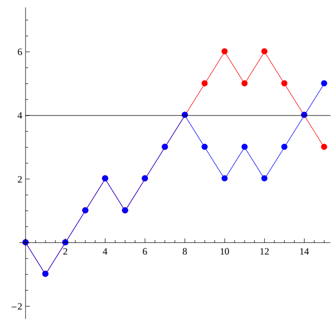

Figure 3. A trajectory (blue) and what you get (red) when you flip its "tail" after the first visit to $l = 4$ which happened at time $F(\omega) = 8$. Note that the red trajectory takes the same values as the blue one up to $n = 8$.

The punchline is that the trajectories in $A_l^>$ as well as in $A_l^=$ are easy to count. For them, the requirement that the level l is hit at a certain point is redundant; if you are at or above l at the very end, you must have hit l at a certain point (if nothing else, at time n). Therefore, $A_l^>$ is simply the family of those ω whose final positions are somewhere strictly above l, and, therefore,

$$\mathbb{P}[A_l^>] = \mathbb{P}[X_n = l+1 \text{ or } X_n = l+2 \text{ or} \dots \text{or } X_n = n] = \sum_{k=l+1}^{n} \mathbb{P}[X_n = k],$$

and, therefore, by the reflection principle,

$$\mathbb{P}[A_l^<] = \mathbb{P}[A_l^>] = \sum_{k=l+1}^{n} \mathbb{P}[X_n = k].$$

We still need to account for $A_l^=$, i.e., for the trajectories that end up exactly at the level l. Just like above,

$$\mathbb{P}[A_l^=] = \mathbb{P}[X_n = l].$$

Putting all of this together, we get

$$\mathbb{P}[A_l] = \mathbb{P}[X_n = l] + 2\sum_{k=l+1}^{n} \mathbb{P}[X_n = k],$$

so that

$$\begin{aligned}\mathbb{P}[M_n = l] &= \mathbb{P}[M_n \geq l] - \mathbb{P}[M_n \geq l+1] = \mathbb{P}[A_l] - \mathbb{P}[A_{l+1}] \\ &= \mathbb{P}[X_n = l] + \mathbb{P}[X_n = l+1]. \qquad \square\end{aligned}$$

To show the versatility of the reflection principle, let us use it to solve a classical problem in combinatorics.

Example 3.5.2 (The Ballot Problem)**.** Suppose that two candidates, Daisy and Oscar, are running for office, and $T \in \mathbb{N}$ voters cast their ballots. Votes are counted the old-fashioned way, namely by the same official, one by one, until all T of them have been processed. After each ballot is opened, the official records the number of votes each candidate has received so far. At the end, the official announces that Daisy has won by a margin of $k > 0$ votes, i.e., that Daisy got $(T+k)/2$ votes and Oscar the remaining $(T-k)/2$ votes. What is the probability that at no time during the counting has Oscar been in the lead?

We assume that the order in which the official counts the votes is completely independent of the actual votes, and that each voter chooses Daisy with probability $p \in (0,1)$ and Oscar with probability $q = 1-p$. For $0 \leq n \leq T$, let X_n be the number of votes received by Daisy *minus* the number of votes received by Oscar in the first n ballots. When the $n+1$-st vote is counted, X_n either increases by 1 (if the vote was for Daisy), or decreases by 1 otherwise. The votes are independent of each other and $X_0 = 0$, so X_n, $0 \leq n \leq T$ is a simple random walk with the time horizon T. The probability of an up-step is $p \in (0,1)$, so this random walk is not necessarily symmetric. The ballot problem can now be restated as follows:

For a simple random walk $\{X_n\}_{0 \leq n \leq T}$, what is the probability that $X_n \geq 0$ for all $n \in \{0, \dots, T\}$, given that $X_T = k$?

The first step towards understanding the solution is the realization that the exact value of p does not matter. Indeed, we are interested in the conditional probability $\mathbb{P}[F|G] = \mathbb{P}[F \cap G]/\mathbb{P}[G]$, where F denotes the set of ω whose corresponding trajectories always stay nonnegative, while the trajectories corresponding to $\omega \in G$ reach k at time n. Each $\omega \in G$ consists of exactly $(T+k)/2$ up-steps (1s) and $(T-k)/2$ down steps (-1s), so its probability weight is equal to $p^{(T+k)/2} q^{(T-k)/2}$. Therefore, with $\#A$ denoting the number of elements in the set A, we get

$$\mathbb{P}[F|G] = \frac{\mathbb{P}[F \cap G]}{\mathbb{P}[G]} = \frac{\#(F \cap G)\ p^{(T+k)/2} q^{(T-k)/2}}{\#G\ p^{(T+k)/2} q^{(T-k)/2}} = \frac{\#(F \cap G)}{\#G}.$$

This is quite amazing in and of itself. This conditional probability does not depend on p at all!

Since we already know how to count the number of elements in G (there are $\binom{T}{(T+k)/2}$), "all" that remains to be done is to count the number of elements in $G \cap F$. The elements in $G \cap F$ form a portion of all the elements in G whose trajectories don't hit the level $l = -1$; this way, $\#(G \cap F) = \#G - \#H$, where H is the set of all paths which finish at k, but cross (or, at least, touch) the level $l = -1$ in the process. Can we use the reflection principle to find $\#H$? Yes, we can. In fact, you can convince yourself that the reflection of any trajectory corresponding to $\omega \in H$ around the level $l = -1$ after its last hitting time of that level produces a trajectory that starts at 0 and ends at $-k-2$, and vice versa. The number of paths from 0 to $-k-2$ is easy to count - it is equal to $\binom{T}{(T+k)/2+1}$. Putting everything together, we get

$$\mathbb{P}[F|G] = \frac{\binom{T}{n_1} - \binom{T}{n_1+1}}{\binom{T}{n_1}} = \frac{k+1}{n_1+1}, \text{ where } n_1 = \frac{T+k}{2}.$$

The last equality follows from the definition of binomial coefficients $\binom{T}{i} = \frac{T!}{i!(T-i)!}$.

The Ballot problem has a long history (going back to at least 1887) and has spurred a lot of research in combinatorics and probability. In fact, people still write research papers on some of its generalizations. When posed outside the context of probability, it is often phrased as "*in how many ways can the counting be performed ...*" (the difference being only in the normalizing factor $\binom{T}{n_1}$ appearing in Example 3.5.2 above). A special case $k = 0$ seems to be even more popular - the number of $2n$-step paths from 0 to 0 never going below zero is called the **Catalan number** and equals to

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{3.5.1}$$

Here is another nice consequence of the reflection principle, i.e., its application to the running maximum. Our formula for the distribution of the maximum of the random walk on $\{0, \dots, T\}$ can be used to answer the following question:

What is the probability that the random walk will reach the level l in T steps (or fewer)?

Indeed, $\{X_n\}_{n\in\mathbb{N}}$ will reach l during the first T steps if and only if $M_T \geq l$. Therefore the answer to the above question is $\mathbb{P}[M_T \geq l] = \mathbb{P}[X_T \geq l] + \mathbb{P}[X_T \geq l+1]$. A special case is the following:

What is the probability that X will stay at or below 0 throughout the interval $\{0, \dots, T\}$?

Clearly, $\{X_n\}_{n\in\mathbb{N}}$ will stay non-positive if it never hits the level 1. The probability of that is $\mathbb{P}[M_T = 0] = \mathbb{P}[X_T = 0] + \mathbb{P}[X_T = 1]$. What happens to this expression as T get larger and larger. In other words, if I give my walk enough time, can I guarantee that it will reach the level 1? Let us compute. For simplicity, let us consider only even time horizons $T = 2N$, so that $\mathbb{P}[M_T = 0] = \mathbb{P}[X_{2N} = 0]$. Using the formula for the distribution of X_{2N}, we get

$$\mathbb{P}[X_{2N} = 0] = \binom{2N}{N} 2^{-2N},$$

so our problem reduces to the investigation of the behavior of $\binom{2N}{N} 2^{-2N}$, as N gets larger and larger, i.e.,

$$\lim_{N\to\infty} \binom{2N}{N} 2^{-2N}. \tag{3.5.2}$$

To evaluate this limit, we need to know about the precise asymptotics of $N!$, as $N \to \infty$:

Proposition 3.5.3 (Stirling's formula)**.** *We have*

$$N! \sim \sqrt{2\pi N} \left(\tfrac{N}{e}\right)^N,$$

where $A_N \sim B_N$ *means* $\lim_{N\to\infty} \frac{A_N}{B_N} = 1$.

Let us use Stirling's formula in (3.5.2):

$$\binom{2N}{N} 2^{-2N} = \frac{(2N)!}{N!N!} 2^{-2N} \sim \frac{\sqrt{2\pi 2N}(2N/e)^{2N}}{\sqrt{2\pi N}(N/e)^N \sqrt{2\pi N}(N/e)^N} 2^{-2N}$$
$$= \tfrac{1}{\sqrt{\pi N}}$$

Therefore,

$$\lim_{N\to\infty} \binom{2N}{N} 2^{-2N} = 0,$$

and it follows that the answer to our question is positive:

Yes, the simple symmetric random walk will reach the level 1*, with certainty, given enough time.*

By symmetry, the level 1 can be replaced by -1. Also, once we hit 1, the random walk "renews itself" (this property is called the Strong Markov Property and we will talk about it later), so it will eventually hit the level 2, as well. Continuing the same way, we get the following remarkable result

> **Theorem 3.5.4.** *The symple symmetric random walk will visit any point in* $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$*, eventually.*

3.6 Problems

Problem 3.6.1. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. The distribution of the product X_1X_2 is

(a) $\begin{array}{c|cc} & 0 & 2 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$

(b) $\begin{array}{c|ccc} & -2 & 0 & 2 \\ \hline & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}$

(c) $\begin{array}{c|ccccc} & -2 & -1 & 0 & 1 & 2 \\ \hline & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{array}$

(d) $\begin{array}{c|ccc} & -1 & 0 & -1 \\ \hline & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}$

(e) none of the above

Problem 3.6.2. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk with the time horizon $T = 3$. The probability that X will never hit the level 2 or the level -2 is

(a) $\frac{1}{4}$ (b) $\frac{1}{3}$ (c) $\frac{1}{2}$ (d) $\frac{3}{8}$ (e) none of the above

Problem 3.6.3. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. Then

(a) X_1 and X_2 are independent

(b) $X_4 - X_2$ is independent of X_3.

(c) $X_4 - X_2$ is independent of $X_6 - X_5$

(d) $X_1 + X_3$ is independent of $X_2 + X_4$

(e) none of the above

Problem 3.6.4. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. Then

(a) X_1 and X_2 are independent

(b) $X_3 - X_1$ is independent of X_2.

(c) $\mathbb{P}[X_{32} = 44 | X_{12} = 0] = \mathbb{P}[X_{22} = 44 | X_2 = 0, X_1 = 1]$

(d) $\mathbb{P}[X_{13} = 4] = \binom{13}{4} 2^{-13}$

(e) none of the above

Problem 3.6.5. Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a simple symmetric random walk. Which of the following processes are simple random walks?

1. $\{2X_n\}_{n \in \mathbb{N}_0}$?
2. $\{X_n^2\}_{n \in \mathbb{N}_0}$?
3. $\{-X_n\}_{n \in \mathbb{N}_0}$?
4. $\{Y_n\}_{n \in \mathbb{N}_0}$, where $Y_n = X_{5+n} - X_5$?

How about the case $p \neq \frac{1}{2}$?

Problem 3.6.6. Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a biased simple random walk with $p = \mathbb{P}[X_1 = 1] = 1/3$. Compute the following:

1. $\mathbb{P}[X_2 = 0]$
2. $\mathbb{P}[X_7 = X_{16}]$
3. $\mathbb{P}[X_2 = X_4 = X_8]$,
4. $\mathbb{P}[X_n < 10 \text{ for all } 0 \leq n \leq 10]$.

Problem 3.6.7. Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a symmetric simple random walk. Compute the following

1. $\mathbb{P}[X_{2n} = 0]$, $n \in \mathbb{N}_0$,
2. $\mathbb{P}[X_n = X_{2n}]$, $n \in \mathbb{N}_0$,
3. $\mathbb{P}[\,|X_1 X_2 X_3| = 2]$,
4. $\mathbb{P}[X_7 + X_{12} = X_1 + X_{16}]$.

Problem 3.6.8. Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a symmetric simple random walk with $p = \mathbb{P}[X_1 = 1] \in (0, 1)$. For $n \in \mathbb{N}$, the probability $\mathbb{P}[X_{2n} = X_{4n} \text{ and } X_{6n} = X_{8n}]$ is given by

(a) $\binom{2n}{n} 2^{-4n}$

(b) $\binom{n}{\lfloor n/2 \rfloor} 2^{-4n}$

(c) $\binom{4n}{2n} 2^{-4n}$

(d) $\binom{2n}{n}^2 2^{-4n}$

(e) none of the above

Problem 3.6.9. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple random walk with $\mathbb{P}[X_1 = 1] = p \in (0,1)$. Define

$$Y_n = \tfrac{1}{n} \sum_{k=1}^{n} X_k, \text{ for } n \in \mathbb{N}.$$

Compute $\mathbb{E}[Y_n]$ and $\operatorname{Var}[Y_n]$, for $n \in \mathbb{N}$.

Hint: You can use the following formulas:

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}, \qquad \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$$

without proof.

Problem 3.6.10. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. Given $n \in \mathbb{N}_0$ and $k \in \mathbb{N}$, compute $\operatorname{Var}[X_n]$, $\operatorname{Cov}[X_n, X_{n+k}]$ and $\operatorname{corr}[X_n, X_{n+k}]$, where Cov stands for the covariance and corr for the correlation. **Note:** If you forgot what these are, look them up.

Problem 3.6.11. Let $\{X_n\}_{0\leq n\leq 10}$ be a simple symmetric random walk with time horizon $T = 10$. What is the probability it will never reach the level 5?

Problem 3.6.12. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. Given $n \in \mathbb{N}$, what is the probability that X does not visit 0 during the time interval $1, \dots, n$.

Problem 3.6.13. Luke starts a random walk, where each step takes him to the left or to the right, with the two alternatives being equally likely and independent of the previous steps. 11 steps to his right is a cookie jar, and Luke gets to take a (single) cookie every time he reaches that position. He performs exactly 15 steps, and then stops.

1. What is the probability that Luke will be exactly by the cookie jar when he stops?

2. What is the probability that Luke stops with with exactly 3 cookies in his hand?

3. What is the probability that Luke stops with at least one cookie in his hand?

4. Suppose now that we place a bowl of broccoli soup one step to the right of the cookie jar. It smells so bad that, if reached, Luke will throw away all the cookies he is currently carrying (if any) and run away pinching his nose. What is the probability that Luke will finish his 15-step walk without ever encountering the yucky bowl of broccoli soup and with at least one cookie in his hand?

Problem 3.6.14. A fair coin is tossed repeatedly and the record of the outcomes is kept. Tossing stops the moment the total number of heads obtained so far exceeds the total number of tails by 3. For example, a possible sequence of tosses could look like *HHTTTHHTHHTHH*. What is the probability that the length of such a sequence is at most 10?

Problem 3.6.15. (*) Let C_n denote the n-th Catalan number (defined in (3.5.1)).

1. Use the reflection principle to show that C_n is the number of trajectories $(x_0, \dots, x_{2n})$ of a random walk with time horizon $T = 2n$ such that $x_k \geq 0$, for all $k \in \{0, 1, \dots, 2n\}$ and $x_{2n} = 0$.

2. Prove the *Segner's recurrence formula* $C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$. (*Hint:* Don't compute - just think about paths).

3. Show that C_n is the number of ways that the vertices of a regular $2n$-gon can be paired so that the line segments joining paired vertices do not intersect.

4. Prove that
$$C_n = \binom{2n}{n} - \binom{2n}{n+1},$$
both algebraically (using the formula for the binomial coefficient) and combinatorially (by counting).