

## 6. The Simple Random Walk

### The Basic Process

Suppose that  $X = (X_1, X_2, \dots)$  is a sequence of independent random variables, each taking values 1 and  $-1$  with probabilities  $p$  and  $1 - p$  respectively. In this section, we will study the **partial sum process**  $Y = (Y_0, Y_1, Y_2, \dots)$  associated with  $X$ . Thus,

$$Y_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N}$$

The sequence  $Y$  is the **simple random walk** with parameter  $p$ . At each step, our random walker moves either one unit to the right (with probability  $p$ ) or one unit to the left (with probability  $1 - p$ ). The walker could accomplish this by tossing a coin with probability of heads  $p$  at each step, to determine whether to move right or move left. Other types of **random walks** are studied in the chapter on **Markov Chains**.

1. Let  $\mu$  and  $\sigma$  denote the **mean** and **standard deviation**, respectively of a step  $X$ . Show that

- $\mu = 2p - 1$
- $\sigma^2 = 4p(1 - p)$

2. Let  $I_j = \frac{1}{2}(X_j + 1)$  for  $j \in \mathbb{N}_+$ . Show that  $I_j = 1$  if  $X_j = 1$  and  $I_j = 0$  if  $X_j = -1$ .

Hence,  $I = (I_1, I_2, \dots)$  is a sequence of **Bernoulli trials**. In terms of the random walker,  $I_j$  is the indicator variable for the event that the  $i^{\text{th}}$  step is to the right.

3. Let  $Z_n = \sum_{j=1}^n I_j$  for  $n \in \mathbb{N}$ .

- Show that  $Y_n = 2Z_n - n$  for  $n \in \mathbb{N}$ .
- Show that  $Z_n$  has the **binomial distribution** with trial parameter  $n$  and success parameter  $p$ .

In terms of the walker,  $Z_n$  is the number of steps to the right in the first  $n$  steps.

4. Use the results of the previous exercise to show that  $Y_n$  has probability density function

$$\mathbb{P}(Y_n = k) = \binom{n}{(n+k)/2} p^{(n+k)/2} (1-p)^{(n-k)/2}, \quad k \in \{-n, -n+2, \dots, n-2, n\}$$

5. Show that

- $\mathbb{E}(Y_n) = n(2p - 1)$
- $\text{var}(Y_n) = 4np(1 - p)$

## The Symmetric Simple Random Walk

Suppose now that  $p = \frac{1}{2}$ . In this case,  $Y = (Y_0, Y_1, \dots)$  is called the **symmetric, simple random walk**. The symmetric random walk can be analyzed using some special and clever combinatorial arguments.

6. Show that the random vector  $X_n = (X_1, X_2, \dots, X_n)$  is **uniformly distributed** on  $S = \{-1, 1\}^n$ , and therefore

$$\mathbb{P}(X_n \in A) = \frac{\#(A)}{2^n}, \quad A \subseteq S$$

7. Show that  $Y_n$  has probability density function

$$\mathbb{P}(Y_n = k) = \binom{n}{(n+k)/2} \frac{1}{2^n}, \quad k \in \{-n, -n+2, \dots, n-2, n\}$$

8. Show that  $Y_n$  has mean and variance given by

- $\mathbb{E}(Y_n) = 0$
- $\text{var}(Y_n) = n$

9. In the **random walk simulation**, select the final position. Vary the number of steps and note the shape and location of the probability density function and the mean/standard deviation bar. For selected values of the parameter, run the simulation 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density and moments to the true density and moments.

10. In the **random walk simulation**, select the final position and set the number of steps to 50. Run the simulation 1000 times, updating every 10 runs. Compute and compare the following:

- $\mathbb{P}(-6 \leq Y_{50} \leq 10)$
- The relative frequency of the event  $\{-6 \leq Y_{50} \leq 10\}$
- The normal approximation to  $\mathbb{P}(-6 \leq Y_{50} \leq 10)$



## The Maximum Position

Consider again the simple, symmetric random walk. Let  $M_n = \max\{Y_0, Y_1, \dots, Y_n\}$ , the **maximum position** during the first  $n$  steps. Note that  $M_n$  takes values in the set  $\{0, 1, \dots, n\}$ . The distribution of  $M_n$  can be derived from a simple and wonderful idea known as the **reflection principle**.

11. Show that  $M_n \geq m$  if and only if  $Y_i = m$  for some  $i \in \{0, 1, \dots, n\}$ .

12. Suppose that  $k \leq m \leq n$ . Show that for each path that satisfies  $M_n \geq m$  and  $Y_n = k$  there is another path that satisfies  $Y_n = 2m - k$ . *Hint:* The second path is obtained from the first path by reflecting in the line  $y = m$ , after the first path hits  $m$ .

13. Use the results of the previous two exercises and the fact that the paths are equally likely to show that

$$\mathbb{P}(M_n \geq m, Y_n = k) = \mathbb{P}(Y_n = 2m - k), \quad k \leq m \leq n$$

14. Use the result of the previous exercise to show that

$$\mathbb{P}(M_n = m, Y_n = k) = \mathbb{P}(Y_n = 2m - k) - \mathbb{P}(Y_n = 2(m + 1) - k), \quad k \leq m \leq n$$

15. Use the result of the previous exercise to show that for  $n \in \mathbb{N}$  and  $m \in \{0, 1, \dots, n\}$

$$\mathbb{P}(M_n = m) = \begin{cases} \mathbb{P}(Y_n = m) = \binom{n}{(m+n)/2} \frac{1}{2^n}, & m \text{ and } n \text{ have the same parity (both even or both odd)} \\ \mathbb{P}(Y_n = m + 1) = \binom{n}{(m+n+1)/2} \frac{1}{2^n}, & m \text{ and } n \text{ have opposite parity (one even and one odd)} \end{cases}$$

16. In the **random walk simulation**, select the maximum value variable. Vary the number of steps and note the shape and location of the density function and the mean/standard deviation bar. Now set the number of steps to 30 and run the simulation 1000 times with an update frequency of 10. Note the apparent convergence of the relative frequency function and empirical moments to the true density function and moments.

17. Show that for any  $n$ , the probability density function of  $M_n$  is decreasing.

The result in Exercise 17 is a bit surprising; in particular, the single most likely value for the maximum is 0!

18. Explicitly compute the probability density function, mean, and standard deviation of  $M_5$ .



19. A fair coin is tossed 10 times. Find the probability that the difference between the number of heads and the number of tails is never greater than 4.



### The Last Visit to 0

Consider again the simple, symmetric random walk. Our next topic is the last visit to 0 during the first  $2n$  steps:

$$L_{2n} = \max \{j \in \{0, 2, \dots, 2n\} : Y_j = 0\}, \quad n \in \mathbb{N}$$

Note that since visits to 0 can only occur at even times,  $L_{2n}$  takes the values in the set  $\{0, 2, \dots, 2n\}$ . This random variable has a strange and interesting distribution known as the **discrete arcsine distribution**. Along the way to our derivation, we will discover some other interesting results as well.

20. Show that

$$\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(Y_{2k} = 0, Y_{2k+1} \neq 0, \dots, Y_{2n} \neq 0), \quad k \in \{0, 1, \dots, n\}$$

21. Use independence, symmetry and the result of the previous exercise to show that

$$\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(Y_{2k} = 0) \mathbb{P}(Y_1 \neq 0, Y_2 \neq 0, \dots, Y_{2n-2k} \neq 0), \quad k \in \{0, 1, \dots, n\}$$

We know the first factor on the right from the distribution of  $Y_{2k}$ . Thus, we need to compute the second factor, the probability that our random walk never returns to 0 during a time interval.

22. Use results for the **maximum position** to show that

$$\mathbb{P}(Y_1 \leq 0, Y_2 \leq 0, \dots, Y_{2j} \leq 0) = \mathbb{P}(M_{2j} = 0) = \binom{2j}{j} \frac{1}{2^{2j}}$$

23. Use symmetry (which is just the reflection principle at  $y = 0$ ), to show that

$$\mathbb{P}(Y_1 \geq 0, Y_2 \geq 0, \dots, Y_{2n} \geq 0) = \binom{2n}{n} \frac{1}{2^{2n}}$$

24. Show that  $\{Y_1 > 0, Y_2 > 0, \dots, Y_{2j} > 0\} = \{Y_1 = 1, Y_2 \geq 1, \dots, Y_{2j} \geq 1\}$

25. Use independence, symmetry, and the result of the previous exercise to show that

$$\mathbb{P}(Y_1 > 0, Y_2 > 0, \dots, Y_{2j} > 0) = \mathbb{P}(Y_1 = 1) \mathbb{P}(Y_1 \geq 0, Y_2 \geq 0, \dots, Y_{2j-1} \geq 0)$$

26. Show that  $Y_{2j-1} \geq 0$  implies  $Y_{2j} \geq 0$ .

27. Use the result of the previous exercises to show that

$$\mathbb{P}(Y_1 > 0, Y_2 > 0, \dots, Y_{2j} > 0) = \binom{2j}{j} \frac{1}{2^{2j+1}}$$

28. Use symmetry and the result of the previous exercise to show that

$$\mathbb{P}(Y_1 \neq 0, Y_2 \neq 0, \dots, Y_{2j} \neq 0) = \binom{2j}{j} \frac{1}{2^{2j}}$$

29. Use the result of [Exercise 22](#) and [Exercise 28](#) to show that the probability density function of  $L_{2n}$  is

$$\mathbb{P}(L_{2n} = 2k) = \binom{2k}{k} \binom{2n-2k}{n-k} \frac{1}{2^{2n}}, \quad k \in \{0, 1, \dots, n\}$$

30. In the [random walk simulation](#), choose the last visit to 0 and then vary the number of steps with the scroll bar. Note the shape and location of the density function and the mean/standard deviation bar. For various values of the parameter, run the simulation 1000 times with an update frequency of 10 and watch the apparent convergence of the empirical density function and moments to the true probability density function and moments.

31. Show that

- $\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(L_{2n} = 2n - 2k)$  so the probability density function is symmetric about  $n$ .
- $\mathbb{P}(L_{2n} = 2j) > \mathbb{P}(L_{2n} = 2k)$  if and only if  $j < k$  and  $2k \leq n$ , so the probability density function is  $u$ -shaped.

In particular, 0 and  $2n$  are the most likely values. The arcsine distribution is quite surprising. Since we are tossing a fair coin to determine the steps of the walker, you might easily think that the random walk should be positive half of the time and negative half of the time, and that it should return to 0 frequently. But in fact, the arcsine law implies that with probability  $\frac{1}{2}$ , there will be *no* return to 0 during the second half of the walk, from time  $n + 1$  to  $2n$ , regardless of  $n$ , and it is not uncommon for the walk to stay positive (or negative) during the entire time from 1 to  $2n$ .

32. Explicitly compute the probability density function, mean, and variance of  $L_{10}$ .



## The Ballot Problem and the First Return to Zero

### The Ballot Problem

Suppose that in an election, candidate  $A$  receives  $a$  votes and candidate  $B$  receives  $b$  votes where  $a > b$ . Assuming a random ordering of the votes, what is the probability that  $A$  is always ahead of  $B$  in the vote count? This is an historically famous problem known as the **Ballot Problem**, that was solved by **Joseph Louis Bertrand** in 1887. The ballot problem is intimately related to simple random walks.

33. Comment on the validity of the assumption that the voters are randomly ordered for a real election.

The ballot problem can be solved by using a simple **conditional probability** argument to obtain a recurrence relation. Let  $f(a, b)$  denote the probability that  $A$  is always ahead of  $B$  in the vote count.

34. Condition on the candidate that receives the last vote to show that

$$f(a, b) = \frac{a}{a+b} f(a-1, b) + \frac{b}{a+b} f(a, b-1)$$

35. Use the initial condition  $f(1, 0) = 1$  and induction on the number of votes  $n = a + b$  to show that

$$f(a, b) = \frac{a-b}{a+b}$$

36. In the **ballot experiment**, vary the parameters  $a$  and  $b$  and note the change the ballot probability. For selected values of the parameters, run the experiment 1000 times with and update frequency of 10 and note the apparent convergence of the relative frequency to the true probability.

37. In an election for mayor of a small town, Mr. Smith received 4352 votes while Ms. Jones received 7543 votes. Compute the probability that Jones was always ahead of Smith in the vote count.



## Relation to Random Walks

Consider again the **simple random walk**  $Y$  with parameter  $p$ .

38. Given  $Y_n = k$ , show that

- There are  $\frac{n+k}{2}$  steps to the right and  $\frac{n-k}{2}$  steps to the left.
- All possible orderings of the steps to the right and the steps to the left are equally likely.

39. Use the previous exercise and the **ballot probability** to show that for  $k > 0$ ,

$$\mathbb{P}(Y_1 > 0, Y_2 > 0, \dots, Y_{n-1} > 0 | Y_n = k) = \frac{k}{n}$$

40. In the **ballot experiment**, vary the parameters  $a$  and  $b$  and note the change the ballot probability. For selected values of the parameters, run the experiment 1000 times with an update frequency of 10 and note the apparent convergence of the relative frequency to the true probability.

41. An American roulette wheel has 38 slots; 18 are red, 18 are black, and 2 are green. Fred bet \$1 on red, at even stakes, 50 times, winning 22 times and losing 28 times. Find the probability that Fred's net fortune was always negative.



Roulette is studied in more detail in the chapter on [Games of Chance](#).

### The Distribution of the First Zero

Consider again the simple random walk with parameter  $p$ , as in the last subsection. Let  $T$  denote the time of the first return to 0:

$$T = \min \{n \in \mathbb{N}_+ : Y_n = 0\}$$

Note that returns to 0 can only occur at even times; it may also be possible that the random walk never returns to 0. Thus,  $T$  takes values in the set  $\{2, 4, \dots\} \cup \{\infty\}$ .

42. Show that

$$\mathbb{P}(T = 2n) = \mathbb{P}(T = 2n, Y_{2n} = 0) = \mathbb{P}(T = 2n | Y_{2n} = 0) \mathbb{P}(Y_{2n} = 0)$$

43. Use the ballot problem to show that

$$\mathbb{P}(T = 2n | Y_{2n} = 0) = \frac{1}{2n-1}$$

44. Use the results of the previous two exercises to show that

$$\mathbb{P}(T = 2n) = \binom{2n}{n} \frac{1}{2n-1} p^n (1-p)^n, \quad n \in \mathbb{N}_+$$

45. Fred and Wilma are tossing a fair coin; Fred gets a point for each head and Wilma gets a point for each tail. Find the probability that their scores are equal for the first time after  $n$  tosses, for each  $n \in \{2, 4, 6, 8, 10\}$ .



[Virtual Laboratories](#) > [11. Bernoulli Trials](#) > [1](#) [2](#) [3](#) [4](#) [5](#) **6**

[Contents](#) | [Applets](#) | [Data Sets](#) | [Biographies](#) | [External Resources](#) | [Keywords](#) | [Feedback](#) | ©