# 11.6: The Simple Random Walk

The simple random walk process is a minor modification of the Bernoulli trials process. Nonetheless, the process has a number of very interesting properties, and so deserves a section of its own. In some respects, it's a discrete time analogue of the Brownian motion process.

## The Basic Process

Suppose that $\boldsymbol{U} = (U_1, U_2, \ldots)$ is a sequence of independent random variables, each taking values 1 and $-1$ with probabilities $p \in [0, 1]$ and $1 - p$ respectively. Let $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ be the partial sum process associated with $\boldsymbol{U}$, so that

$$X_n = \sum_{i=1}^{n} U_i, \quad n \in \mathbb{N} \tag{11.6.1}$$

The sequence $\boldsymbol{X}$ is the *simple random walk* with parameter $p$.

We imagine a person or a particle on an axis, so that at each discrete time step, the walker moves either one unit to the right (with probability $p$) or one unit to the left (with probability $1 - p$), independently from step to step. 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, and additional properties of this random walk, are studied in the chapter on Markov Chains.

The mean and standard deviation, respectively of a step $U$ are

1. $\mathbb{E}(U) = 2p - 1$
2. $\operatorname{var}(U) = 4p(1 - p)$

Let $I_j = \frac{1}{2}(U_j + 1)$ for $j \in \mathbb{N}_+$. Then $\boldsymbol{I} = (I_1, I_2, \ldots)$ is a Bernoulli trials sequence with success parameter $p$.

Proof

Note that $I_j = 1$ if $U_j = 1$ and $I_j = 0$ if $I_j = -1$.

In terms of the random walker, $I_j$ is the indicator variable for the event that the $j$th step is to the right.

Let $R_n = \sum_{i=1}^{n} I_j$ for $n \in \mathbb{N}$, so that $\boldsymbol{R} = (R_0, R_1, \ldots)$ is the partial sum process associated with $\boldsymbol{I}$. Then

1. $X_n = 2R_n - n$ for $n \in \mathbb{N}$.
2. $R_n$ has the binomial distribution with trial parameter $n$ and success parameter $p$.

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

$X_n$ has probability density function

$$\mathbb{P}(X_n = k) = \binom{n}{(n+k)/2} p^{(n+k)/2}(1-p)^{(n-k)/2}, \quad k \in \{-n, -n+2, \ldots, n-2, n\} \tag{11.6.2}$$

Proof

Since $R_n$ takes values in $\{0, 1, \ldots, n\}$, $X_n$ takes values in $\{-n, -n+2, \ldots, n-2, n\}$. For $k$ in this set, $\mathbb{P}(X_n = k) = \mathbb{P}[R_n = (n+k)/2]$, so the result follows from the binomial distribution of $R_n$.

The mean and variance of $X_n$ are

1. $\mathbb{E}(X_n) = n(2p - 1)$
2. $\operatorname{var}(X_n) = 4np(1 - p)$

## The Simple Symmetric Random Walk

Suppose now that $p = \frac{1}{2}$. In this case, $\boldsymbol{X} = (X_0, X_1, \ldots)$ is called the *simple symmetric random walk*. The symmetric random walk can be analyzed using some special and clever combinatorial arguments. But first we give the basic results above for this special case.

For each $n \in \mathbb{N}_+$, the random vector $\boldsymbol{U}_n = (U_1, U_2, \ldots, U_n)$ is uniformly distributed on $\{-1, 1\}^n$, and therefore

$$\mathbb{P}(\boldsymbol{U}_n \in A) = \frac{\#(A)}{2^n}, \quad A \subseteq S \tag{11.6.3}$$

$X_n$ has probability density function

$$\mathbb{P}(X_n = k) = \binom{n}{(n+k)/2} \frac{1}{2^n}, \quad k \in \{-n, -n+2, \ldots, n-2, n\} \tag{11.6.4}$$

The mean and variance of $X_n$ are

1. $\mathbb{E}(X_n) = 0$
2. $\operatorname{var}(X_n) = n$

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$\pm$standard deviation bar. For selected values of the parameter, run the simulation 1000 times and compare the empirical density function and moments to the true probability density function and moments.

In the random walk simulation, select the final position and set the number of steps to 50. Run the simulation 1000 times and compute and compare the following:

1. $\mathbb{P}(-6 \le X_{50} \le 10)$
2. The relative frequency of the event $\{-6 \le X_{50} \le 10\}$
3. The normal approximation to $\mathbb{P}(-6 \le X_{50} \le 10)$

Answer

1. 0.7794
3. 0.7752

### The Maximum Position

Consider again the simple, symmetric random walk. Let $Y_n = \max\{X_0, X_1, \ldots, X_n\}$, the *maximum position* during the first $n$ steps. Note that $Y_n$ takes values in the set $\{0, 1, \ldots, n\}$. The distribution of $Y_n$ can be derived from a simple and wonderful idea known as the *reflection principle*.

For $n \in \mathbb{N}$ and $y \in \{0, 1, \ldots, n\}$,

$$\mathbb{P}(Y_n = y) = \begin{cases} \mathbb{P}(X_n = y) = \binom{n}{(y+n)/2} \frac{1}{2^n}; & y, n \text{ have the same parity (both even or both odd)} \\ \mathbb{P}(X_n = y+1) = \binom{n}{(y+n+1)/2} \frac{1}{2^n}; & y, n \text{ have opposite parity (one even and one odd)} \end{cases} \tag{11.6.5}$$

Proof

Note first that $Y_n \ge y$ if and only if $X_i = y$ for some $i \in \{0, 1, \ldots, n\}$. Suppose that $k \le y \le n$. For each path that satisfies $Y_n \ge y$ and $X_n = k$ there is another path that satisfies $X_n = 2y - k$. The second path is obtained from the first path by reflecting in the line $x = y$, after the first path hits $y$. Since the paths are equally likely,

$$\mathbb{P}(Y_n \ge y, X_n = k) = \mathbb{P}(X_n = 2y - k), \quad k \le y \le n \tag{11.6.6}$$

Hence it follows that

$$\mathbb{P}(Y_n = y, X_n = k) = \mathbb{P}(X_n = 2y - k) - \mathbb{P}[X_n = 2(y+1) - k], \quad k \le y \le n \tag{11.6.7}$$

In the random walk simulation, select the maximum value variable. Vary the number of steps and note the shape and location of the probability density function and the mean/standard deviation bar. Now set the number of steps to 30 and run the simulation 1000 times. Compare the relative frequency function and empirical moments to the true probability density function and moments.

For every $n$, the probability density function of $Y_n$ is decreasing.

The last result is a bit surprising; in particular, the single most likely value for the maximum (and hence the mode of the distribution) is 0.

Explicitly compute the probability density function, mean, and standard deviation of $Y_5$.

Answer

1. Probability density function of $Y_5$: $f(0) = f(1) = \frac{10}{32}$ , $f(2) = f(3) = \frac{5}{32}$ , $f(4) = f(5) = \frac{1}{32}$
2. $\mathbb{E}(Y_5) = \frac{11}{8}$
3. $\text{var}(Y_5) = \frac{111}{64}$

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.

Answer

$\mathbb{P}(Y_{10} \le 4) = \frac{57}{64}$

## 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:

$$Z_{2n} = \max\{j \in \{0, 2, \ldots, 2n\} : X_j = 0\}, \quad n \in \mathbb{N} \tag{11.6.8}$$

Note that since visits to 0 can only occur at even times, $Z_{2n}$ takes the values in the set $\{0, 2, \ldots, 2n\}$. 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.

The probability density function of $Z_{2n}$ is

$$\mathbb{P}(Z_{2n} = 2k) = \binom{2k}{k}\binom{2n-2k}{n-k}\frac{1}{2^{2n}}, \quad k \in \{0, 1, \ldots, n\} \tag{11.6.9}$$

Proof

Note that

$$\mathbb{P}(Z_{2n} = 2k) = \mathbb{P}(X_{2k} = 0, X_{2k+1} \ne 0, \ldots, X_{2n} \ne 0), \quad k \in \{0, 1, \ldots, n\} \tag{11.6.10}$$

From independence and symmetry it follows that

$$\mathbb{P}(Z_{2n} = 2k) = \mathbb{P}(X_{2k} = 0)\mathbb{P}(X_1 \ne 0, X_2 \ne 0, \ldots, X_{2n-2k} \ne 0), \quad k \in \{0, 1, \ldots, n\} \tag{11.6.11}$$

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

$$\mathbb{P}(X_1 \le 0, X_2 \le 0, \ldots, X_{2j} \le 0) = \mathbb{P}(Y_{2j} = 0)\binom{2j}{j}\frac{1}{2^{2j}} \tag{11.6.12}$$

From symmetry (which is just the reflection principle at $y = 0$), it follows that

$$\mathbb{P}(X_1 \ge 0, X_2 \ge 0, \ldots, X_{2n} \ge 0) = \binom{2n}{n}\frac{1}{2^{2n}} \tag{11.6.13}$$

Next, $\{X_1 > 0, X_2 > 0, \ldots, X_{2j} > 0\} = \{X_1 = 1, X_2 \ge 1, \ldots, X_{2j} \ge 1\}$ . From independence and symmetry,

$$\mathbb{P}(X_1 > 0, X_2 > 0, \ldots, X_{2j} > 0\} = \mathbb{P}(X_1 = 0)\mathbb{P}(X_1 \ge 0, X_2 \ge 0, \ldots, X_{2j-1} \ge 0) \tag{11.6.14}$$

But $X_{2j-1} \ge 0$ implies $X_{2j} \ge 0$. Hence

$$\mathbb{P}(X_1 > 0, X_2 > 0, \ldots, X_{2j} > 0) = \binom{2j}{j}\frac{1}{2^{2j+1}} \tag{11.6.15}$$

From symmetry,

$$\mathbb{P}(X_1 \ne 0, X_2 \ne 0, \ldots, X_{2j} \ne 0) = \binom{2j}{j}\frac{1}{2^{2j}} \tag{11.6.16}$$

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 probability density function and the mean/standard deviation bar. For various values of the parameter, run the simulation 1000 times and compare the empirical density function and moments to the true probability density function and moments.

The probability density function of $Z_{2n}$ is symmetric about $n$ and is $u$-shaped:

1. $\mathbb{P}(Z_{2n} = 2k) = \mathbb{P}(Z_{2n} = 2n - 2k)$
2. $\mathbb{P}(Z_{2n} = 2j) > \mathbb{P}(Z_{2n} = 2k)$ if and only if $j < k$ and $2k \le n$

In particular, 0 and $2n$ are the most likely values and hence are the modes of the distribution. The discrete 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 $2n$.

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

Answer

1. Probability density function of $Z_{10}$: $f(0) = f(10) = \frac{63}{256}$, $f(2) = f(8) = \frac{35}{256}$, $f(4) = f(6) = \frac{30}{256}$
2. $\mathbb{E}(Z_{10}) = 5$
3. $\text{var}(Z_{10}) = 15$

## 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.

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.

$f$ satisfies the initial condition $f(1, 0) = 1$ and the following recurrence relation:

$$f(a, b) = \frac{a}{a+b} f(a-1, b) + \frac{b}{a+b} f(a, b-1) \tag{11.6.17}$$

Proof

This follows by conditioning on the candidate that receives the last vote.

The probability that $A$ is always ahead in the vote count is

$$f(a, b) = \frac{a - b}{a + b} \tag{11.6.18}$$

Proof

This follows from the recurrence relation and induction on the total number of votes $n = a + b$

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 and compare the relative frequency to the true probability.

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.

Answer

$\frac{3191}{11895} \approx 0.2683$

## Relation to Random Walks

Consider again the simple random walk $\boldsymbol{X}$ with parameter $p$.

Given $X_n = k$,

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

For $k > 0$,

$$\mathbb{P}(X_1 > 0, X_2 > 0, \ldots, X_{n-1} > 0 \mid X_n = k) = \frac{k}{n} \tag{11.6.19}$$

Proof

This follows from the previous result and the ballot probability.

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 and compare the relative frequency to the true probability.

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.

Answer

$\frac{3}{25}$

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:

$$T = \min\{n \in \mathbb{N}_+ : X_n = 0\} \tag{11.6.20}$$

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 $\{2, 4, \ldots\} \cup \{\infty\}$.

The probability density funtion of $T_{2n}$ is given by

$$\mathbb{P}(T = 2n) = \binom{2n}{n} \frac{1}{2n-1} p^n (1-p)^n, \quad n \in \mathbb{N}_+ \tag{11.6.21}$$

Proof

For $n \in \mathbb{N}_+$

$$\mathbb{P}(T = 2n) = \mathbb{P}(T = 2n, X_{2n} = 0) = \mathbb{P}(T = 2n \mid X_{2n} = 0)\mathbb{P}(X_{2n} = 0) \tag{11.6.22}$$

From the ballot problem,

$$\mathbb{P}(T = 2n \mid X_{2n} = 0) = \frac{1}{2n-1} \tag{11.6.23}$$

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 $n \in \{2, 4, 6, 8, 10\}$.

Answer

$f(2) = \frac{1}{2}$, $f(4) = \frac{1}{8}$, $f(6) = \frac{1}{16}$, $f(8) = \frac{5}{128}$, $f(10) = \frac{7}{512}$