Random Walks

Alex Boden

HL IB Math

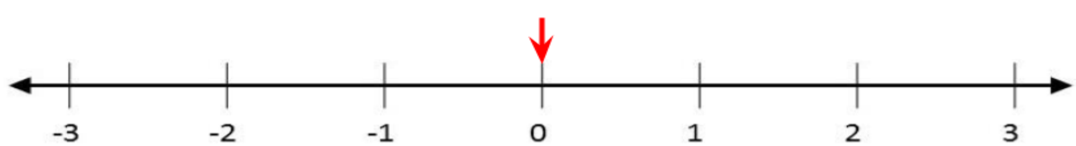
Internal Assessment

February 2nd, 2021

The stock market and its predictability recently captured my interest. After taking a course in building financial security at my school I was inspired to read multiple books on the subject. I have come to the conclusion that day trading on the stock market is essentially gambling. Day trading is when a stock trader attempts to profit for the intraday movements of the stock market. I believe that nobody can consistently profit from the short-term movements of the stock markets as they are seemingly random. Burton Malkiel, the author of "A Random Walk Down Wall Street", corroborates this with his famous quote, "*A blindfolded monkey throwing darts at a newspaper's financial pages could select a portfolio that would do just as well as one carefully selected by experts.*"

This concept of randomness is one that can be expressed mathematically. One of the ways that stochastic (randomly determined) processes are expressed is as a random walk. A random walk consists of a succession of random steps that can be described as a path in some mathematical space. Random walks have many real-world applications such as in topics like computer science, physics, chemistry, biology and economics. **In this internal assessment my aim is to simulate random walks using a computer program and assess how accurate the simulation's results are by comparing them to mathematically derived values.**

The most rudimentary example of is a random walk in one dimension. A one-dimensional random walk can be visualized by placing a marker on a number line at the number zero and allowing it to take n random steps in either direction. Each step moves the marker one unit to the left or right with equal probability.



When we take n steps, the number of different combinations is 2^n because for every step there are two possibilities ($+1$ or -1) meaning that after each step the number of combinations double. In this case each combination is equally likely although some combinations will yield the same ending position for the marker. For example, if $n = 3$ we can have $(+1, +1, -1)$, $(+1, -1, +1)$ and $(-1, +1, +1)$, which will all end up with the marker at 1.

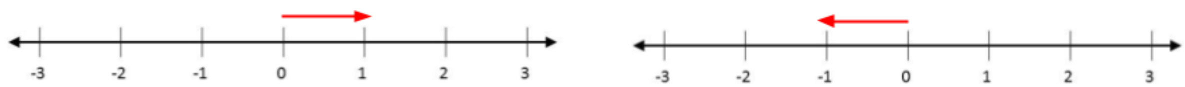
To simulate a one-dimensional random walk, I wrote a simple computer program that randomly chooses +1 or -1 and iterates n times. The displacement of the random walk is denoted D_n and is equal to the sum of these n numbers. In order to find an approximation of D_n, my program simulated a large number of random walks for each value of n and then calculated the mean of these values. The source code for the program is provided below and was written and executed using Python 3.

```
from random import *

highest_n_value = 10
simulations = 1000000000
for n in range(highest_n_value):
    average = 0.0
    for i in range(simulations):
        position = 0.0
        for j in range(n):
            x = randint(1,2)
            if (x==1):
                position+=1
            else:
                position -=1
        average += abs(position)

    print("Average for {}: {}".format(n,average/simulations))
```

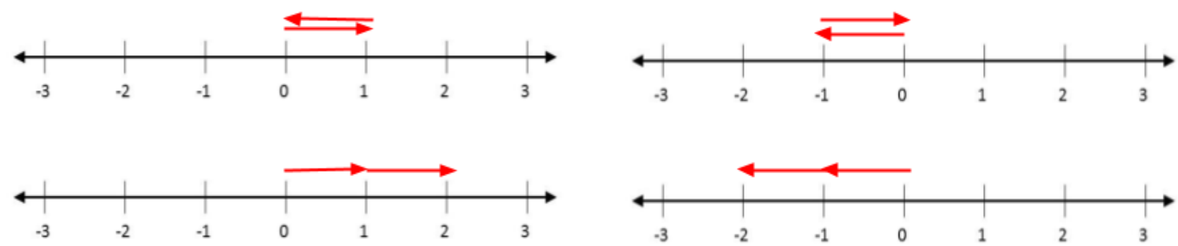
The table below shows the output of the program from n = 1 to n = 10.

Number of steps, $\mathbf{n}$	**Number of Simulations**	**Average distance from the origin, $\boldsymbol{D}_\mathbf{n}$**
1	1 000 000 000	1.0
2	1 000 000 000	0.999995714
3	1 000 000 000	1.499975204
4	1 000 000 000	1.500052974
5	1 000 000 000	1.874965946
6	1 000 000 000	1.875013834
7	1 000 000 000	2.187536476
8	1 000 000 000	2.187481924
9	1 000 000 000	2.461045992
10	1 000 000 000	2.460905928

In the table I noticed that the average distance only seems to increase after every two and that the average distance for $2n$ steps and $2n - 1$ steps are nearly equal. If we take the simplest example of n = 1 and n = 2 we can see why this is true. For n = 1 there are 2^1 possible combinations. They are shown below.



In both cases the displacement from the origin is one, therefore no matter which way it moves after 1 move it will always be 1 unit away. This explains why for n = 1 we got exactly 1. As for n = 2 there are 2^2or 4 combinations which are shown below.



Out of the 4 combinations 2 outcomes yield a distance of 2 and the other 2 outcomes yield a distance of 0 making the average distance one. The reason that the program did not get exactly 1 and instead got 0.999995714 is because it came up with the combinations that yield 0 more often than the ones that yield 2, this is purely random.

Now that we have the data from our simulation, we want to assess its accuracy. To do this we need to calculate the expected value of D_{n}. To help us calculate the expected value of D_n we can use an interesting property I discovered during my research: the binomial coefficients give the numbers of occurrences for a given displacement. The binomial coefficients are numbers that follow Pascal's Triangle.

End position on the number line

	-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
$\boldsymbol{n}$ = 0								1							
$\boldsymbol{n}$ = 1							1		1						
$\boldsymbol{n}$ = 2						1		2		1					
$\boldsymbol{n}$ = 3					1		3		3		1				
$\boldsymbol{n}$ = 4				1		4		6		4		1			
$\boldsymbol{n}$ = 5			1		5		10		10		5		1		
$\boldsymbol{n}$ = 6		1		6		15		20		15		6		1	
$\boldsymbol{n}$ = 7	1		7		21		35		35		21		7		1

This chart has the ending position along the top and increasing n values along the side. These values happen to produce Pascal's Triangle. Coming back to our example where n = 2 we know there are four combinations: one that lands on -2, one that lands on 2 and two that land on 0. In our chart I have highlighted the n = 2 row and we can see that these numbers line up with the example done earlier.

Pascal's Triangle is defined such that the *n*th row and *k*th column are $\binom{\mathrm{n}}{\mathrm{k}}$ where $\binom{\mathrm{n}}{\mathrm{k}} = \frac{n!}{k!(n-k)!}$. We can evaluate some examples for n = 2, which have been highlighted red in the diagram. In Pascal's Triangle the rows and columns start at 0 such that the top 1 in the triangle is $\binom{0}{0}$

$\binom{2}{0}$ = 1, This represents the number of ways we can end on -2 with 2 steps

$\binom{2}{1}$ = 2, This represents the number of ways we can end on 0 with 2 steps

$\binom{2}{2}$ = 1, This represents the number of ways we can end on 2 with 2 steps

Note the symmetry in Pascal's Triangle as it will be important to calculating the expected value of D_{n}.

If we would like to end up at 0, we can calculate the number of combinations that land at 0 for after n steps. To land on 0, the number of steps we take to the right must be equal to the number of steps to the left. The only way for this to happen is if $n = 2m$ is an even number, and if we take m steps to the right and m steps to the left, where $m = n/2$. Following this logic, in order to land up at a positive integer k on the number line, we must take k more steps to the right than steps to the left. Likewise, to land at a negative integer $k = -|k|$ on the number line, we must take $|k|$ more steps to the left than steps to the right. One easy way to think about this is by thinking in terms of a coin flip where heads represents +1 and tails represents -1. To solve an equation for the number of heads we need to end on a given k we can start by letting h be the number of heads. Tails will occur whenever heads do not thus, we can define tails as $n - h$. Then k represents the ending position of the random walk and is equal to the number of heads minus the number of tails. Therefore,

$$k \;=\; h - (n - h) = 2h - n$$

and we see that

$$h = \frac{k \;+\; n}{2}$$

Therefore, heads must appear $\frac{k+n}{2}$ times in order to end on an integer k on the number line. We can incorporate the number of different ways to achieve the result of k with the properties from Pascal's Triangle to define the probability of arriving at k as $\frac{1}{2^n}\binom{n}{(k+n)/2}$.

We can use this information to develop an equation for the expected or theoretical value of the average distance from the starting point. Since each combination is equally likely this will be the sum of the displacements over the number of combinations. We can represent it with the following equation:

$$E(D_n) = \frac{1}{2^n}\sum_{k\,=\,-n}^{n}\binom{n}{(k+n)/2}|\mathrm{k}| \text{ where } (k+n)\; mod\; 2 = 0$$

Where:

- n is the number of steps
- k represents where the marker lands on the number line. We take the absolute value to get the displacement
- $E(D_{\mathrm{n}})$ is the expected value of for displacement after n steps
- $\sum_{k\,=\,-n}^{n}$ We are summing from -n to n as these are the limits of where the marker can land
- $(k+n)\; mod\; 2 = 0$ this means when $(k+n)$ is divided by 2 the remainder is 0. In other words, $(k+n)$ is even. This is an important condition because you cannot land on an odd number by taking an even number of steps or vice versa.
- $\binom{n}{(k+n)/2}$ this gives us the number of times the marker will land on k
- $\frac{1}{2^n}$ We have to divide by the number of combinations to get the average

We can simplify this because the sum is symmetrical, the values for k and $-\mathrm{k}$ are essentially counted twice as we noticed from Pascal's Triangle earlier. Since our sum is over positive integers, we can replace $|\mathrm{k}|$ with k:

$$E(D_{\mathrm{n}}) = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{\frac{(k+n)}{2}}2k \text{ where } (k+n)\ mod\ 2 = 0$$

There is a problem with our current formula because it sums when $\frac{(k+n)}{2}$ is not an integer. Currently we have implemented a conditional statement that $(k+n)\ mod\ 2 = 0.$ To fix this we will have to manipulate the equation such that the bottom term of the $\binom{n}{k}$ is an integer. We can start to simplify our equation by letting $j = n + k$

$$E(D_{\mathrm{n}}) = \frac{1}{2^n}\sum_{j=n}^{2n}\binom{n}{\frac{\mathrm{j}}{2}}2(\mathrm{j}-\mathrm{n}) \text{ where } j\ mod\ 2 = 0$$

Factoring out the 2 we get

$$E(D_{\mathrm{n}}) = \frac{1}{2^{n-1}}\sum_{j=n}^{2n}\binom{n}{\frac{\mathrm{j}}{2}}(\mathrm{j}-\mathrm{n}) \text{ where } j\ mod\ 2 = 0$$

Letting $j = 2i$

$$E(D_{\mathrm{n}}) = \frac{1}{2^{n-1}}\sum_{i=\left\lceil\frac{n}{2}\right\rceil}^{n}\binom{n}{i}(2\mathrm{i}-\mathrm{n})$$

To get rid of our conditional statement we added the ceiling ($\left\lceil\frac{x}{y}\right\rceil$) operation. This rounds up to the nearest integer. For example: $\lceil 2.5 \rceil = 3$. Using this simplified formula, we can compute the average distance for n steps. The results are shown below:

Number of steps, n	$E(D_{\mathrm{n}})$	**Number of steps, n**	$E(D_{\mathrm{n}})$
0	0	5	15/8
1	1	6	15/8
2	1	7	35/16
3	3/2	8	35/16
4	3/2	9	315/128

We see again that the values for $E(\mathrm{D}_{2\mathrm{m}})$ and $E(\mathrm{D}_{2\mathrm{m}-1})$ are the same. We can compare these calculated values to the ones generated before by the computer program. We will round each to 3 significant figures.

Number of steps, n	D_{n} from computer program	$D_{\mathrm{n}}$from calculations	Percentage Difference
1	1.00	1.00	0%
2	1.00	1.00	0%
3	1.50	1.50	0%
4	1.50	1.50	0%
5	1.87	1.88	0.532%
6	1.88	1.88	0%
7	2.19	2.19	0%
8	2.19	2.19	0%
9	2.46	2.46	0%
10	2.46	2.46	0%

When rounded to 3 significant figures the results are all the same with the expectation of when n = 5. Our expected value was $\frac{15}{8}$ or 1.875. This number is equally close to 1.87 and 1.88 but by convention we round up. The average value our program calculated for n = 5 was 1.874965946 which is extremely close to 1.875. This difference in our computer program's value and calculated value is due to a unique rounding circumstance as if we had rounded to 4 significant figures, they would be the same. Therefore, we know that our computer program was highly accurate.

We have seen in both our program and calculations from our formula that the expected displacement values for an even number and the preceding odd number are the same. There is a simple way to think about why this is true. When we have a "walker" or "marker" at some point on the number line d after taking a step it will either be one unit closer to the origin or one unit farther. The average of these outcomes would bring us back to the same point we started with. For example, if you are at d you can go to $d+1$ or $d-1$ and the average of $d+1$ and $d-1$ is d. If this were true for every integer our average distance would remain unchanging, but it is not. The one integer this rule does not hold true for is 0. After taking one step from 0 the average distance is 1, we know this from our earlier calculations of $E(S_1)=1$. This means the average distance will only increase when going up from an even number of steps to an odd number of steps as it takes an even number of steps to end on 0. This explains why $E(\mathrm{D}_{2\mathrm{m}})$ = $E(\mathrm{D}_{2\mathrm{m}-1})$. We can prove why $E(\mathrm{D}_{2\mathrm{m}})$ = $E(\mathrm{D}_{2\mathrm{m}-1})$ using our equation.

This proof will involve the use of Pascal's Rule which states:

$$\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$$

To prove Pascal's Rule, we can rewrite the left-hand side of the equation using the definition of binomial expansion,

$$\binom{n}{r-1} + \binom{n}{r} = \frac{n!}{(r-1)!\,\big(n-(r-1)\big)!} + \frac{n!}{(r)!\,(n-r)!}$$

$$= \frac{n!}{(r-1)!\,\big(n-(r-1)\big)!} \times \frac{r}{r} + \frac{n!}{(r)!\,(n-r)!} \times \frac{n-r+1}{n-r+1}$$

$$= \frac{rn!}{(r)!\,(n-r+1)!} + \frac{n!\,(n-r+1)}{(r)!\,(n-r+1)!} = \frac{rn! + n!\,(n-r+1)}{(r)!\,(n-r+1)!}$$

$$= \frac{n!\,(r+n-r+1)}{(r)!\,(n-r+1)!} = \frac{(n+1)!}{(r)!\,((n+1)-r)!} = \binom{n+1}{r} \blacksquare$$

Taking our original formula

$$E(D_{\mathrm{n}}) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{\frac{(k+n)}{2}} 2k \text{ where } (k+n) \bmod 2 = 0$$

We will have to use a clever trick by substituting $(n-k)$ for $(n+k)$. We can do this because of the symmetry of the Pascal's Triangle and Binomial Expansion. Similarly, to above, to prove this symmetry we can be in the two values and expand using the definition of a binomial expansion:

$$\binom{n}{\frac{(n+k)}{2}} = \frac{n!}{\left(\frac{(n+k)}{2}\right)!\left(n-\left(\frac{(n+k)}{2}\right)\right)!} = \frac{n!}{\left(\frac{(n+k)}{2}\right)!\left(\frac{(n-k)}{2}\right)!}$$

$$\binom{n}{\frac{(n-k)}{2}} = \frac{n!}{\left(\frac{(n-k)}{2}\right)!\left(n-\left(\frac{(n-k)}{2}\right)\right)!} = \frac{n!}{\left(\frac{(n-k)}{2}\right)!\left(\frac{(n+k)}{2}\right)!}$$

$$\binom{n}{\frac{(n+k)}{2}} = \binom{n}{\frac{(n-k)}{2}} \blacksquare$$

Therefore, we take k steps away from the centre of Pascal's triangle in either direction we will achieve the same result. Mathematically expressed, we can substitute

$$\binom{n}{\frac{(n-k)}{2}}\ for \binom{n}{\frac{(n\ +k)}{2}}$$

$$E(D_{\mathrm{n}}) = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{\frac{(n-k\ )}{2}}2\mathrm{k}\ \text{ where } (n-\ k)\ mod\ 2 = 0$$

When $n\ =\ 2m$

$$E(D_{2\mathrm{m}}) = \frac{1}{2^{2m}}\sum_{k=0}^{2m}\binom{2\mathrm{m}}{\frac{(2m\ -\ k)}{2}}2\mathrm{k}\ \text{ where } (2m-\ k)\ mod\ 2 = 0$$

Letting $k\ =\ 2a$

$$E(D_{2\mathrm{m}}) = \frac{1}{2^{2m}}\sum_{a=0}^{m}\binom{2m}{m-a}4\mathrm{a}$$

Using Pascal's Rule

$$E(D_{2\mathrm{m}}) = \frac{1}{2^{2m}}\sum_{a=0}^{m}\left[\binom{2m-1}{m-a}+\binom{2m-1}{m-a-1}\right]4\mathrm{a}$$

$$E(D_{2\mathrm{m}}) = \frac{1}{2^{2m}}\sum_{a=0}^{m}\binom{2m-1}{m-a}4\mathrm{a} + \frac{1}{2^{2m}}\sum_{a=0}^{m}\binom{2m-1}{m-a-1}4\mathrm{a}$$

Letting $b\ =\ a\ +\ 1$ on the second term on the right-hand side, we get

$$E(D_{2\mathrm{m}}) = \frac{1}{2^{2m}}\sum_{a=0}^{m}\binom{2m-1}{m-a}4\mathrm{a} + \frac{1}{2^{2m}}\sum_{b=1}^{m}\binom{2m-1}{m-b}4(\mathrm{b}-1)$$

Factoring out the powers of 2 and rewriting the index of the first sum to begin at 1 instead of 0, we get

$$\boldsymbol{E(D_{2\mathrm{m}}) = \frac{1}{2^{2m-2}}\sum_{a=1}^{m}\binom{2m-1}{m-a}\mathrm{a} + \frac{1}{2^{2m-2}}\sum_{b=1}^{m}\binom{2m-1}{m-b}(\mathrm{b}-\ 1)}$$

When $n = 2m - 1$

$$E(D_{2\mathrm{m}-1}) = \frac{1}{2^{2m-1}} \sum_{k=0}^{m} \binom{2m-1}{\frac{(2m-1-k)}{2}} 2\mathrm{k} \text{ where } (k + 2m - 1)\ mod\ 2 = 0$$

Letting $k = 2b - 1$

$$E(D_{2\mathrm{m}-1}) = \frac{1}{2^{2m-1}} \sum_{b=1}^{m} \binom{2m-1}{m-b} 2(2\mathrm{b} - 1)$$

$$E(D_{2\mathrm{m}-1}) = \frac{1}{2^{2m-1}} \sum_{b=1}^{m} \binom{2m-1}{m-b} (2\mathrm{b} + 2\mathrm{b} - 2)$$

Expanding the sum

$$E(D_{2\mathrm{m}-1}) = \frac{1}{2^{2m-1}} \sum_{b=1}^{m} \binom{2m-1}{m-b} (2\mathrm{b}) + \frac{1}{2^{2m-1}} \sum_{b=1}^{m} \binom{2m-1}{m-b} (2b - 2)$$

Factoring out powers of 2

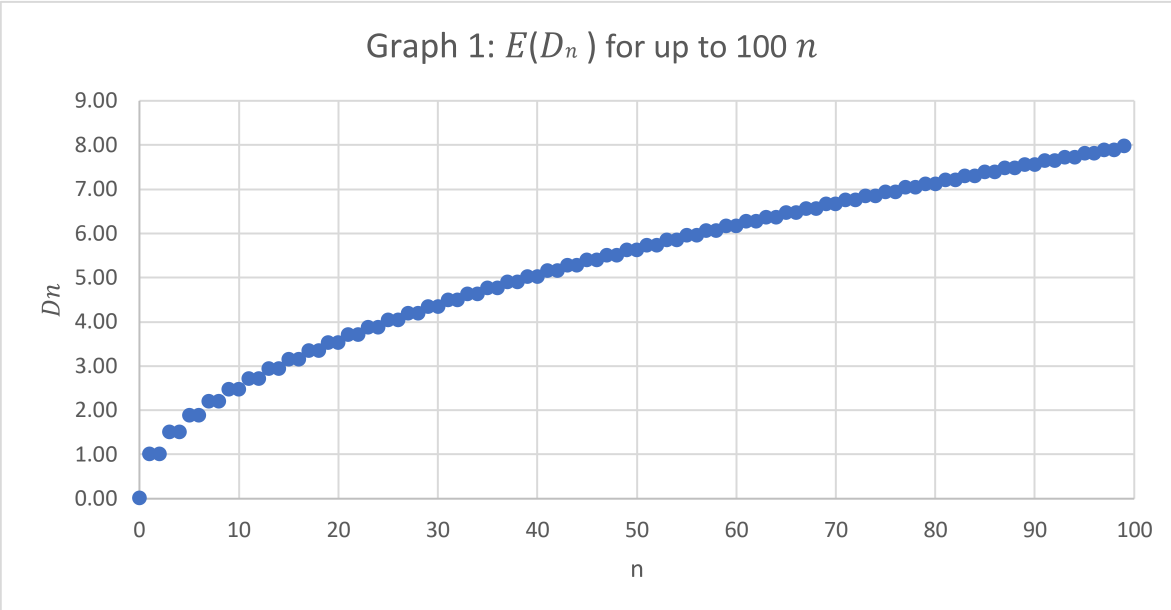
$$\boldsymbol{E(D_{2\mathrm{m}-1}) = \frac{1}{2^{2m-2}} \sum_{b=1}^{m} \binom{2m-1}{m-b} \mathrm{b} + \frac{1}{2^{2m-2}} \sum_{b=1}^{m} \binom{2m-1}{m-b} (\mathrm{b} - 1)}$$

$$\boldsymbol{E(D_{2\mathrm{m}}) = E(D_{2\mathrm{m}-1})} \blacksquare$$

Now that we understand how to calculate the $E(D_n)$ for a given n we can write a program that calculates the $E(D_n)$ for higher values as calculations by hand would be lengthy and tedious. Using Python 3, we can output the results for any value of n. This program will output the first 100 values for $E(D_n)$:

```
from fractions import Fraction

def fac(x):
    result = 1
    for i in range(1,x+1):
        result *= i
    return result

def bionom(n, k):
    return fac(n)/(fac(n-k)*fac(k))

highest_n_value = 100
for n in range(highest_n_value):
    dn = 0
    for k in range(1, n+1):
        if((n+k)%2 == 0):
            dn += bionom(n,(n+k)/2)*2*k
    print('{}'.format(Fraction(dn, 2**n)))
```

Plotting the first 100 values in Excel, we end up with graph 1



Graph 1 resembles a $f(x) = \sqrt{x}$ function. Upon further investigation the points do not exactly follow any $f(x) = a\sqrt{x}$ model even when every other point is taken out. Note there can be no vertical or horizontal shift as the first point is $[0,0]$.

We have seen and proved that $E(D_{2m})$ and $E(D_{2m-1})$ and equal. We relate $E(D_{2m})$ and $E(D_{2m+1})$ by the number of walkers on 0 as we have already established these are only factor effecting the $E(D_n)$. From Pascal's Triangle above we know that when a $n = 2\text{m}$ we can calculate the number of walkers on 0 with the formula $\binom{n}{\frac{(n-k)}{2}}$. Since $\text{k} = 0$ using the formula for $n = 2\text{m}$ we get $\binom{2m}{m}$. Since every walker on 0 will end up on 1 or -1 the average displacement for these walkers will be 1. Following this we can relate $E(D_{2m})$ and $E(D_{2m+1})$ as:

$$E(D_{2m+1}) = \frac{E(D_{2m}) \times 2^{2m} + \binom{2m}{m}}{2^{2m}}$$

Or

$$E(D_{2m+1}) = E(D_{2m}) + \frac{\binom{2m}{m}}{2^{2m}}$$

We can write a Python 3 program to calculate these values recursively:

```
from fractions import Fraction

def fac(x):
    result = 1
    for i in range(1,x+1):
        result *= i
    return result

def bionom(n, k):
    return fac(n)/(fac(n-k)*fac(k))

total = 0

for n in range(8):
    if(n % 2 == 0):
        total += float(bionom(n,n/2))/float((2**n))
        print(Fraction(total))
```

Output:

Number of steps, n	$\boldsymbol{E(D_n)}$
1	1
3	3/2
5	15/8
7	35/16

The output consists of every other value since we know $E(D_{2m})$ is equal to $E(D_{2m-1})$.

Calculating the $E(D_n)$ recursively using this gives us the same values as we calculated mathematically. This further increases the certainty of our results.

Although simple at the surface level one dimensional walks and their expected displacement have some interesting math behind them. In this internal assessment I used computer programs as a tool to investigate the average displacement in one dimensional random walks. The outputs of the initial program that simulated a random walk with n steps gave us insight and into the property that $E(D_{2m}) = E(D_{2m-1})$. Using the formula developed we were able to prove why this is true by using Pascal's Rule and properties of summations.

In conclusion using a computer program to simulate a stochastic process can be highly accurate when a sufficient number of trials are run. I was able to verify the results from the computer program by developing a formula for average displacement. I compared the results from the formula to the values generated by the computer and they were nearly identical when rounded to 3 significant digits. These results were corroborated by those obtained from the recursive formula.

Bibliography

Malkiel, B. G. (2020). *A random walk down Wall Street: The time-tested strategy for successful investing*. New York: W.W. Norton and Company.

Weisstein, E. W. (n.d.). Binomial distribution. Retrieved February 02, 2021, from https://mathworld.wolfram.com/BinomialDistribution.html

Weisstein, E. W. (n.d.). Random walk--1-dimensional. Retrieved February 02, 2021, from https://mathworld.wolfram.com/RandomWalk1-Dimensional.html