Anthony Mazy

Paul Horn, Math 3080

Random Walks: Computation and Visualization

The purpose of this project was to apply the topics we learned in class to Random Walks while introducing an element of experimental probability utilizing data science in python. I wanted to create simulations that would be useful in educational settings as well as for data science. This is an important for me as the whole reason I became a math Major is to give a level of theoretical and mathematical rigor to the data-science/statistics stuff I envision doing for my career.

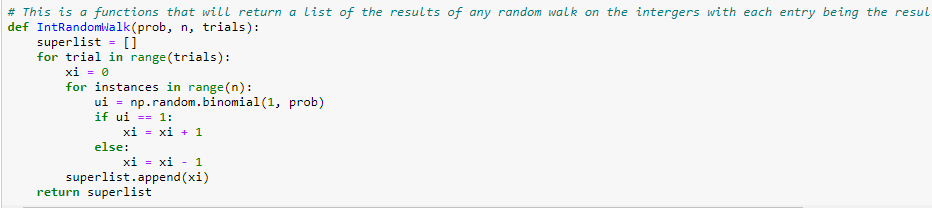
Final Note On Submission: All my simulations are documented in my workbook, which I have uploaded in its original form and as an HTML file

**Random Walks on the Integers:**

U

Xn is our ‘Random Walk’

We will Take a theoretical approach as well as an experimental approach to understand the nature of the random walk on the integers. The first tool we have generated is a function in Jupyter Notebook which returns a certain number of trials of Random Walks on the integers with a given value of p



General observations:

The second key observation is that distribution of the random walk can be related to the binomial distribution because you are essentially subtracting the failures from the number of ‘successes’ (where takes the value 1). Below we can see how the Probability Mass Function of the binomial distribution ‘Maps’ to the pmf of the random walk on the integers

Let Yn Bin(n, p)

(Yn = 0)

(Yn = 1)

(Yn = 2)

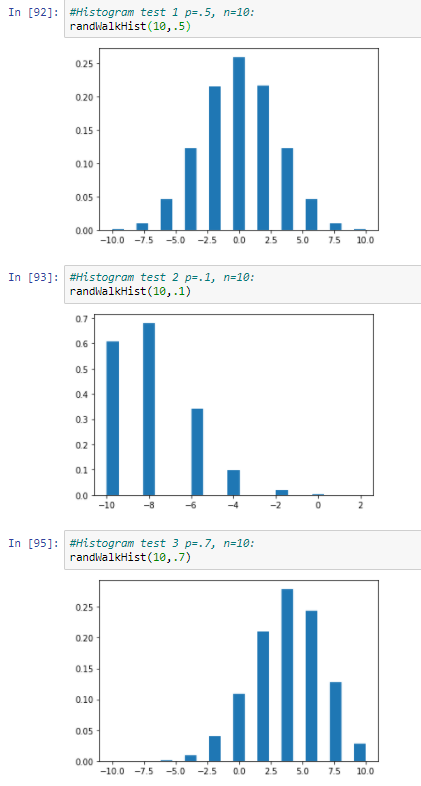
…

(Yn = n)

We can deduce the following formula for the Probability Mass Function of Xn via modifying the binomial PMF

where

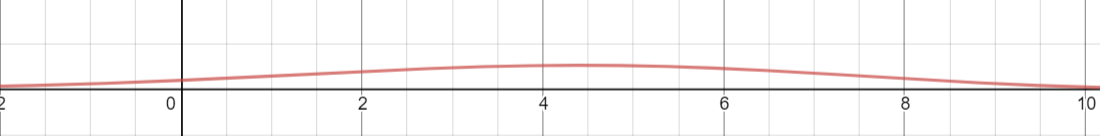
Since the mathematical basis of this experiment is non-existent and instead is an observation is based on an intuition, we can run high trial simulations of random walks and compare them against our hypothesized distribution shape.



These simulations align with our proposed pdfs which are shown below for our chosen cases.





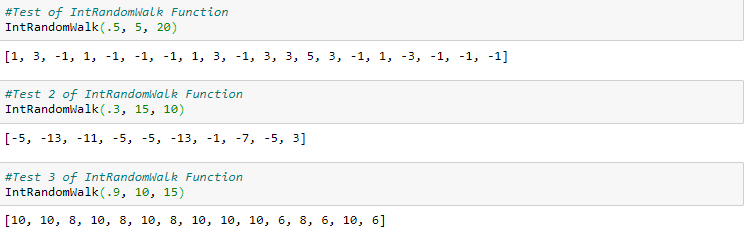


Note: recall our constraint on k

We can also make the following observation regarding the sample space of

note: that is odd when n is odd and is even when n is even

We can view this experimentally with our function in Jupyter Notebook to assess if our statement about holds true



We can see from these simulations where n = 5, 15, and 10 respectively our proposition holds.

**Base Case: Standard Random Walks**

The standard random walk where p = .5 => 1 – p = .5: (i.e. you have the same chance of going forwards or backwards)

let us denote this random walk as Xn

Proposition:

This result makes perfect intuitive sense. If we are equally likely to go forward and backwards on we expect to end up the same place we ended up.

Now let us find the Variance of Xn

Since Ui takes the value -1 and 1 with the equal frequency of 0.5, and we have assessed the expected have we can assess that the standard deviation, σ = 1

Thus, since Var(Ui) = σ2 = 1

thus Var(Xn) = n and Std(Xn) =

Alternatively, we can use the following formula to derive the same response:

thus,

**General Case: where p is not strictly defined**

Now let us consider the case of a random walk where p, the probability that any Ui takes the value 1, is not strictly defined and instead is defined as:

0 ≤ p ≤ 1

With the implication that the probability any Ui takes the value -1 being 1-p

Now let us calculate the expectation of this general case:

We can write this out the better visualize this

∵ All elements in U are independent and have the same distribution with the following Probability Mass Function

Thus, we can say that all elements of U have the same expected value

So,

where P(k) is the probability takes the value k

Now, given the PMF that we already found for we can assess the expectation of

Thus,

Notion of ‘Margins’: to better understand the mechanism of the distribution

In this instance the notion of margins refers to the net expectation of any as defined previously with the 0 ≤ p ≤ 1. As such, our margin per event is 2p – 1, which we can also think of as p – (1-p).

Let margin per event: Mp = 2p -1

For example,

take p = .5. Intuitively and computationally, the expectation of = 0 regardless of n. This is a function of margin. Let margin be M.5

M.5 = (2)(.5) – 1= 0

The interpretation is that after each event, we expect to not move at all. Since all the events in U are uniform and independent, we expect to move nowhere after n moves, thus

Now take p = 1, are certain to move forward after each event thus our Margin is one, so

This note on margins seems trivial at this point, however its importance as a concept shows itself further in this paper where we discuss the dynamic random walk on the integers (where p is not uniform across all the events in the walk)

In addition, let us find the variance of this general case

Var( =

Var( =

Var(

Var(

, where j

Thus,

Var(

This leads us to an interesting remark regarding the variance of , which is that the

Proposition: , a function of the product of p and the complement of p, is maximized where p = .5 and and is minimized where p = 0 or 1 and

Additionally, is increasing on the interval [0,0.5) and decreasing on the interval (0.5,1]

Proof

Computationally, we can prove this using the First Derivative Test

Let be a positive integer, which is one of the fundamental assumptions of a random walk

And let

where p = .5

Thus, there is a critical point at p = 0.5

Now we must test the intervals of p, [0,0.5) and (0.5,1] to determine whether is increasing or decreasing on said intervals

Select p = 0 and p = 1 to test [0,0.5) and (0.5,1] respectively

And

Thus, is increasing on the interval p: [0,0.5) and decreasing on p: (0.5,1] and = n at p = 0.5

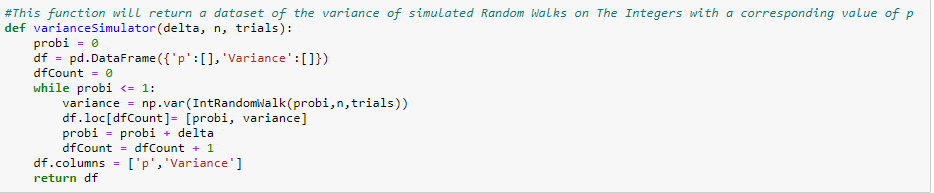
Additionally, since p = 0,1 are endpoints of the interval of valid values of p: [0,1]

= 0 at p = 0, 1

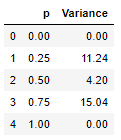
Thus Proposition Proven

We can also take a graphical/experimental approach to understand and visualize the variance:

In JupyterNotebooks, I created the following simulation module that finds the variance of a simulated random walks at various values of p

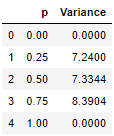


This was the result of the first simulation I ran as a test of the function: with 10 moves in each walk and 10 trials



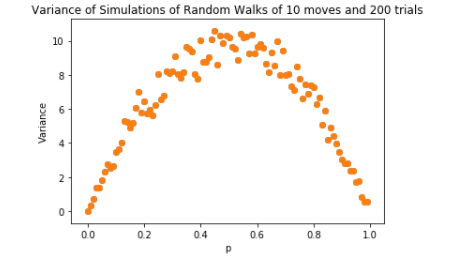
Though this simulation was only meant to test that my model was working, the result gave me pause because I was expecting to see a unimodal distribution of Var(Xn) at p = .5. Instead, I saw a bimodal distribution around p = .25 and .75. In response I ran a second larger scale simulation to verify that the theoretical/computational basis of my work was compatible with the model.

n = 10, trials = 50



This slightly larger scale simulation gave me assurance that my model was properly built and that my computational work did not have a massive error and that the result of the previous simulation was due to a small sample size and random chance.

However, I order to truly understand the behavior of Var(Xn) we need a simulation that takes the variance at more values of p and with more trials. In addition, we’d like to visualize this data to see if it fits our calculations for variance.



The graph above, which shows the variance of a 200 trial simulation of 10 move (n = 10) random walks at various equidistant values of p, shows experimental results that align with our theoretical understanding of the random walk. The graph has a symmetric, unimodal, parabolic that we would expect given our formula for Variance

Var(

Given in this instance n = 200,

Var(

The behavior of which aligns with our Experimental results. The Variance begins at zero increasing rapidly at first to a plateau of approximately 10 at p = 0.5 trailing off again symmetrically to zero at p = 1

**Random Walks on Multiple Dimensions:**

Just a few notes on Random walks on As we have described the notion of random walks on the integers, we can expand this notion of random walks to , , and

**Random Walk on :**

Such That

U

And

V

Tying back to course work, we can state is a joint distribution of independent random variables. As such we are can work out the following joint pmf of the function, however, since the random walk is discrete we write the joint pmf as a double sum not double integral

We can also introduce the notion of magnitude of , defined as such

This notion of distance and magnitude has interesting results and application (I would assume).