# Introduction to the Random Walk

Varshaa Venkitesh

August 2025

## 1 Introduction

It's midnight; the moon is a thin crescent overhead. The roads are empty. A drunkard stumbles out onto the sidewalk, face illuminated only by the glow of the streetlamp, and starts walking home. In his intoxicated state, he is equally as likely to step forwards as he is backwards. After he takes a step, he is again equally as likely to step in either direction.

This "drunkard's walk" is a visualization of the concept of the random walk. It is a process in which a sequence of steps are taken in random directions. In the simplest scenario, a random walker starts at a certain point and takes steps in either of two directions with equal probability.

## 2 Modeling the One-Dimensional Random Walk

For a one-dimensional random walk, the simplest way to define the walk would be to take a certain number (n) of steps and each step is either forward or backward. Mathematically, the walk is defined as a sequence of $n$ independent integers $a_n$, each having equal probability of being either 1 or -1. Then, the random walk can be defined as the series $S_n$:

$$S_n = \sum_{i=0}^{n} a_n \tag{1}$$

If you take a specific value for n and then do many such random walks, the final positions (values) of the walkers tends towards the same result despite each walker taking a unique route. This statistical property is fundamentally due to the central limit theorem, which states that the distribution of sample means is always approximately normal if the sample size is large enough.

This is an interesting property, but for modeling physical systems we need to relate each individual walker to something we can actually observe. One thing we can consider is how far each walker gets after n steps. This can be represented by the sum of all the steps.

$$S_n = a_0 + a_1 + a_2 + ... + a_n \tag{2}$$

If we average all the sums, we would get the expected displacement from the start after taking n steps.

$$\langle S_n \rangle = \langle \sum_{i=0}^{n} a_i \rangle = \sum_{i=0}^{n} \langle a_i \rangle \tag{3}$$

However, the average over many trials, $\langle a_i \rangle$, will just be 0 so the right-hand side of the above equation will be 0. (Even though each individual trial can take a person as far as n units forward or backward, the most probable distance traveled will be 0; right back to the start.) Therefore, the mean displacement will not be useful to us when thinking about the process leading to that end.

What about the squared distance traveled instead? Since both $(-1)^2$ and $1^2$ equal 1, this seems like it would be more useful since the square of the displacement must be nonnegative and the square of any step will be 1.

The sum is still the same: $a_0 + a_1 + a_2 + ... + a_n$ So if we square it: $(a_0 + a_1 + a_2 ...) * (a_0 + a_1 + a_2 ...)$ There will be two kinds of terms– The positive products with squares: $a_0a_0 + a_1a_1 + a_2a_2...$ And the cross-products which could be either negative or positive: $a_0a_1 + a_0a_2 + ... + a_1a_0 + a_1a_2 + ...$

We know that the sum of the cross-products will average out to 0 (similar idea to our previous section with the mean displacement). The positive squares will sum to n. Therefore,

$$\langle S_n^2 \rangle = \sum_{i=0}^{n} \langle a_i^2 \rangle + \sum_{i=0}^{n} \sum_{j \neq i} 2 \langle a_i a_j \rangle = n + 0 = n \tag{4}$$

So, we have a nonzero mean squared displacement (MSD). This means that on average, the walker went a distance of n while on average their displacement from the start was 0 because they go either forward or backward randomly. They just follow a jittery path to get back to where they started.

## 3 The Gaussian

Let's revisit the normal distribution we encountered earlier. The distribution is centered around zero displacement, and looks like an approximately smooth "bell-shaped" curve. If we take the limit as the number of trials approaches infinity, the distribution of results would be a normal distribution given by the expression:

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-((x-\mu)/2\sigma)^2} \tag{5}$$

The mean of the displacement $x$ is $\mu$; for the one-dimensional random walk, this is 0. A curve of the function $e^{-(x/a)^2}$ is called a *Gaussian curve*. The one in this case is centered at 0 and spreads out on either side of the center. For large x, the function goes to 0, and at $x = a$ it goes to $\frac{1}{e}$. We said the curve spreads out on either side; what is its width? The width of the curve is given by the standard deviation $\sigma$.

For the one-dimensional random walk, the standard deviation (with a step size of 1 unit; i.e. a step is either +1 or -1) is:

$$\sigma = \sqrt{n} \tag{6}$$

(How did we arrive at this result? The standard deviation $\sigma$ is the square root of the variance $\sigma^2$. The variance is defined as the average of the squared differences from the mean. After calculating the mean, for each number $a_i$ subtract the mean from $a_i$ and square the result (the squaring is done to ensure a positive number). Then divide the sum of the squared differences by the number of numbers to find the average of the squared differences. In this case, the mean was 0 and each squared result was 1. $1 + 1 + ...$ n times, divided by n steps in a walk, means the variance here was 1. The square root of 1, i.e. the standard deviation, is then 1.)

## 4 Random Walks in Multiple Dimensions

Here, we've covered the one-dimensional random walk in some detail. We can have random walks in multiple dimensions as well. At each step, we would decide in which dimension to move and whether we move forward (positive direction) or backward (negative direction).

# 5 History and Brownian Motion

The most familiar random walk in physics came from a classical problem. In 1827, Scottish botanist Robert Brown was observing pollen grains suspended in water through his microscope. He noticed that the pollen grains kept going in some jittery motion he couldn't explain.

It wasn't until 1905 that Einstein came up with a satisfactory answer to this observation. He used some statistical methods to solve the problem by predicting that the random motion of the pollen grains was because of water being made up of molecules, which were constantly bumping into the pollen grains. Therefore, Brownian motion could be defined as the limit of random walks as the time intervals between steps become infinitesimal. Einstein derived a diffusion equation for the pollen grains and then showed how to link it to quantities such as the MSD.

(The concept of atoms and molecules was not fully accepted at the time, and so Einstein viewed this model of Brownian motion as a way to either prove or disprove the atomic theory. A few years later, Jean Perrin experimentally verified the theory, and the idea of atoms and molecules became standard.)

# 6 Diffusion Equation

Einstein modeled the random motion using a probability function $\rho(x,t)$ ($\rho$ is density), to describe the likelihood of finding a grain at position x at time t. He derived this diffusion equation:

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2} \tag{7}$$

Here, $D$ is the diffusion coefficient, which describes how quickly the grains spread out (how "eager" the grains are to spread out).

(To understand what this equation is actually saying, think of $\rho(x,t)$ as a heat map showing where the grains are likely to be. $\frac{\partial \rho}{\partial t}$ describes how the heat map changes over time. $\frac{\partial^2 \rho}{\partial x^2}$ measures how steep the hills and valleys are in the heat map; i.e. how uneven the distribution is. And, just as a reminder, $D$ describes how quickly the grains spread out. Therefore, the equation is stating that the change in grain distributions over time depends on how uneven the distribution is right now and how fast the grains tend to move. In other words, over time, the more uneven the distribution is, the more it will spread out. And the larger the diffusion coefficient is, the faster that "spreading out" will occur.)

Afterwards Einstein connected this probabilistic description to the MSD, which tells us how far on average a grain moves from the start over time. For one-dimensional diffusion, he showed that:

$$\langle x^2(t) \rangle = 2Dt \tag{8}$$

where $\langle x^2(t) \rangle$ is the MSD at time $t$. Similarly, for the general case:

$$\langle S^2(t) \rangle = 2dDt \tag{9}$$

where $d$ is the number of dimensions in the problem.

This equation is saying, "On average, the square of how far a particle has wandered from its starting point grows linearly with time." As it is a linear relationship, when graphing the MSD vs. the number of steps taken we expect the slope of this line to be the diffusion coefficient.

# 7 Introduction to Applications

Random walks are useful in numerous fields and contexts! Here are just a few of the many ways they are used:

- Physics:
  - Diffusion processes and molecular movement
  - Quantum mechanics; interference and superposition
  - Computational physics across many domains
- Biology:
  - Genetic drift
  - Spread of diseases and bacteria
  - Modeling behavior of animals, cells, etc.
  - Neuroscience and decision-making systems
- Finance:
  - Modeling financial markets
- Computer Science:
  - Monte Carlo simulations
  - Randomized algorithms

## 8 Next Steps

In this paper, we've covered the basic derivation and ideas of the random walk and diffusion equation. In the future, one enhancement to this paper might be to code a few snippets to help visualize properties of the random walk listed in the paper. For example, a snippet could be coded to generate Brownian motion, plot the MSD vs. number of steps taken, or show how after many random walk trials the results tend to a normal distribution. Another future step would be to expand the previous section to include more examples of applications. Furthermore, the applications listed could be explored in greater depth through subsequent papers.

## 9 Resources

I have linked a list of online resources I used when writing this paper. These can also be used for further explanation of concepts discussed as well as visuals and simulations of the ideas presented.

Standard Deviation and Variance:

`https://www.mathsisfun.com/data/standard-deviation.html`

Random Walk:

`https://www.physics.ucla.edu/~chester/TECH/RandomWalk/`

Random Walk with Visuals:

`https://prancer.physics.louisville.edu/modules/random_walks/index.html`

`https://compphys.notes.dmaitre.phyip3.dur.ac.uk/lectures/lecture-8/random-walks/`

`https://web.mit.edu/8.334/www/grades/projects/projects17/OscarMickelin/brownian.html`