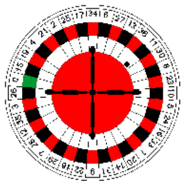

Random Walks

A Random Walk (Real World Phenomena)

Imagine a perfume bottle opened in the front of a classroom and the fragrance soon drifting throughout the room. The fragrance spreads because some molecules evaporate from the bottle and then collide randomly with other molecules in the air, eventually reaching your nose even though you are hidden in the last row. We wish to develop a model for this process which we can then use as the basis for a computer simulation of a random walk. Once we have tested the simulation, we can virtually "see" what the random walk taken by a perfume molecule looks like, and be able to predict the distance the perfume's fragrance travels as a function of time. The model we shall develop to describe the path traveled by a molecule is called a *random walk*. "Random" because it is chance collisions that determine the direction in which a perfume molecule travels, and "walk" because it takes a series of "steps" for the molecule to get from here to there. The same model has been used to simulate the search path of a foraging animal, the fortune of a gambler, and the accumulation of error in computer calculations, among others [1-9].

It is probably true that not all processes modeled as a random walk are truly random, in the sense that random means there is no way of predicting the next step from knowledge of the previous step. However, all these processes do contain an element of *chance* (what mathematicians call *stochastic processes*), and so a random walk, which has chance as a key element, is at least a good starting model for them.

An Aside on Root Mean Square Averages

Before we develop our computer model of a random walk, which is quite simple, we need to develop a simple mathematical model. Then we can use this mathematical model to see if the computer simulation is behaving in a way that we would expect nature's random walk to behave.

The math problem is to predict how many collisions, on the average, a perfume molecule makes in traveling a distance R. You are given the fact that, on the average, a molecule travels a distance r

between collisions. (The velocities of the molecules and distance traveled between collisions increases with temperature, but we shall assume that the temperature is constant.)

Before we can proceed we need to be a bit more precise about what we mean by *average.* In common usage, we may use the word "average" to denote what in statistics is called the *mean*. For example, let's say *N* students take an exam, and the grade for student *i* on the exam is x_i. Then the mean or average score for the exam is

(1) $$\bar{x} = < x > = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sometimes the numbers we want to take the average of can have negative as well as positive values (we hope your test scores do not have the former!). While it is perfectly acceptable to take the average of positive and negative numbers, sometimes we are more interested in what might be called the average *size* of the numbers, irrespective of their sign. In that case we would use another type of average known as the *root mean square* or *rms* average. In an rms average we first find the mean of the square of the numbers

(2) $$\overline{x^2} = < x^2 > = \frac{1}{N}\sum_{i=1}^{N} x^2.$$

Since only positive numbers are being averaged in Equation (2), there is no cancellation of terms in $\overline{x^2}$, but it is a distance squared and not a distance. To obtain a measure of the distance, we take the square root of $\overline{x^2}$,

(3) $$x_{rms} = \sqrt{\overline{x^2}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x^2}.$$

This quantity, the root mean square average of *x,* serves as a measure of the average magnitude of *x*, regardless of what sign *x* may take. Equation (3) is not as bad as it looks. It just means take the average of x^2, and then take the square root of the average.

Random Walk Mathematical Model

Many areas of science make use of a mathematical model of a random walk that predicts the average distance traveled in a walk of *N* steps. In order to verify the validity of our simulated random walk, we will compare the mathematical and simulated results. The mathematical model tells us that the average distance *R* from the origin (see Figure 1 Left) at which the walker ends up after *N* steps of length 1, is equal to $\sqrt{N}$. Of course if you add up the lengths of each step the total distance traveled is *N*, but since the steps are not all in a straight line, we can't just add up their lengths to calculate the distance from the starting point.

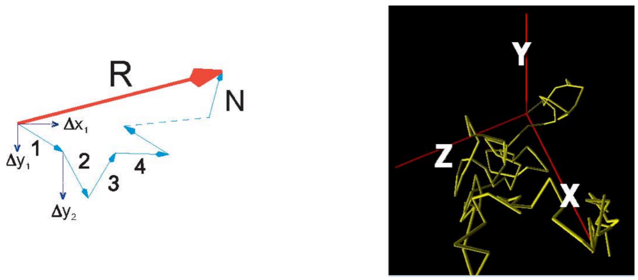

Figure 1. *Left:* A schematic of the N steps taken in a random walk that ends at a distance R from the origin. *Right:* The actual steps taken in a simulation of a 3-D random walk.

Here we present a simple model for a random walk in two dimensions, that is, on a flat surface. The same basic model can be applied to walks in one or three dimensions, or be extended with more sophisticated methods of including chance.

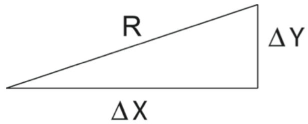

Figure 2. A right triangle with sides equal to the horizontal and vertical components of the length *R* of a random walk.

We assume that a "walker" takes sequential steps, with the direction of each step independent of the direction of the previous step. As seen in Figures 1 and 2, the walker starts at the origin and take *N* steps in the *x*-*y* plane, each of length 1. The first step has a horizontal (*x*) component of Δx_1, and a vertical (*y*) component of Δy_1. The second step has a horizontal component Δx_2, and a vertical component Δy_2, while the last step has a horizontal component Δx_N, and a vertical component Δy_N. Although each step may be in a different direction, the distances along the *x* and *y* axes just add algebraically. Accordingly, ΔX, the total *x* distance from the origin is

(4) $$\Delta X = \Delta x_1 + \Delta x_2 + \ldots + \Delta x_N .$$

Likewise, ΔY, the total y distance is

(5) $$\Delta Y = \Delta y_1 + \Delta y_2 + \ldots + \Delta y_N .$$

The radial distance *R* from the starting point after *N* steps (Figure 1 Left) is the hypotenuse of a right triangle with sides of ΔX and ΔY (Figure 2). We use Pythagoras's theorem to find *R*:

(6) $$R(N)^2 = \Delta X^2 + \Delta Y^2$$

(7) $$R(N) = \sqrt{\Delta X^2 + \Delta Y^2},$$

where the argument *N* is included to indicate that *R* is a function of the number of steps *N.*

Equation (6) gives us the distance traveled in a random walk of *N* steps. Since random walks have chance entering at each step, it is likely that different walks of *N* steps will result in different values for the length *R*. However, we expect that if we *average* over many walks, all with the same number *N* of steps, then we should obtain an average in which some of the random fluctuations have been removed.

So how do we go about taking an average value for *R?* First let's make it clear that we are keeping the number of steps *N* in a walk constant while we take the average. To be explicit, let's say that we average over *M* different walks, all with *N* steps. Since the steps in our random walk are just as likely to go to the right as to the left, or to go up as often as down, if the number of walks *M* is large, then the average over all the walks of $\Delta X(N)$ = 0. Likewise, the average of ΔY would also vanish. Well, that won't do! The walker does end up some finite distance from the starting point every time, and we should be able to make at least an approximate prediction of that distance.

This then is where our old friend, the root-mean-square average of Equation (3) comes to our rescue. Since R^2 is always positive, we can take its average over *M* walks, and then take the square root of that average to get a measure of the average distance from the origin after *N* steps:

$$(8) \qquad \overline{R^2}(N) = \frac{1}{M}\sum_{j=1}^{M}(\overline{R^2})_j$$

$$(9) \qquad R_{rms}(N) = \sqrt{\overline{R^2}(N)}$$

where $\overline{R^2}_j$ indicates one of the j^{th}calculated value of the average $\overline{R^2}$.

If this seems a little confusing, do not worry too much about it now because we'll come back to when we evaluate Equation (8) as part of our simulation. Now we have to do a little algebra, which may seem complicated at first, but then becomes rather simple. Let's take our expression Equation (6) for R^2 and substitute Equations (4) and (5) for ΔX and ΔY:

$$(10) \qquad R(N)^2 = (\Delta X_1 + \Delta X_2 + \ldots + \Delta X_N)^2 +$$

$$(11) \qquad = (\Delta X_1)^2 + (\Delta X_2)^2 + \ldots + (\Delta X_N)^2 + \Delta X_1 \Delta X_2 + \ldots + \Delta X_1 \Delta X_N$$
$$+ (\Delta Y_1)^2 + (\Delta Y_2)^2 + \ldots + (\Delta Y_N)^2 + \Delta Y_1 \Delta Y_2 + \ldots + \Delta Y_1 \Delta Y_N.$$

Now Equation (11) is unquestionably rather messy looking, but not to worry. If we take the average of $R(N)^2$ for a large number *M* of different walks (all with *N* steps*),* then all of the cross terms like $\Delta Y_1\, \Delta Y_2$ will vanish (or average out to a small number) since they are all just as likely to be negative as positive. In contrast, the squared terms like $(\Delta X_1)^2$ are always positive and so do not vanish when we take the average. We are thus left with a much simpler approximation for the average value of $R(N)^2$:

$$(12) \qquad \overline{R^2}(N) \approx (\Delta X_1)^2 + (\Delta X_2)^2 + \ldots + (\Delta X_N)^2 + (\Delta Y_1)^2 + (\Delta Y_2)^2 + \ldots + (\Delta Y_N)^2$$

$$(13) \qquad = [(\Delta X_1)^2 + (\Delta Y_1)^2] + [(\Delta X_2)^2 + (\Delta Y_2)^2] + \cdots + [(\Delta X_N)^2 + (\Delta Y_N)^2].$$

But note, each of the sums in (12) is just the length of the corresponding step in the random walk, which we have already said equals 1. So we have *N* terms each equal to 1, and when we add them all up we get the simple result

$$(14) \qquad \overline{R^2}(N) \approx N.$$

Equation (13) states that the average distance squared after a random walk of *N* steps of length 1 is *N.* If we take the square root of both sides of Equation (13) we obtain the desired expression for the root-mean-square, or rms, radius:

$$(15) \qquad R_{rms}(N) = \sqrt{\overline{R^2}(N)} \approx \sqrt{N}.$$

This is the simple result that characterizes a random walk. To summarize, if the walk is random, then we expect that on the average the walker is just as likely to be on the left as on the right, or as likely to be up as down, or in other words,

$$(16) \qquad \overline{\Delta X} \approx \overline{\Delta Y} \approx 0.$$

However, even though all directions are equally likely, the more steps that the walker takes, the further it gets from the origin, with the rms distance from the origin after *N* steps growing like the square root of *N.* In practice, we expect simulations to agree with equation (14) only when the number of steps is large and only when the number of paths used to take the average is large. Enough talk, let's simulate a random walk and see what we get.