www.iop.org/journals/physed

Randomness at the root of things 1: Random walks

Jon Ogborn[1], **Simon Collins**[2] **and Mick Brown**[3]

[1] Institute of Education, University of London, UK
[2] Stowe School, Buckingham, UK
[3] Cavendish Laboratory, University of Cambridge, UK

Abstract

This is the first of a pair of articles about randomness in physics. In this article, we use some variations on the idea of a 'random walk' to consider first the path of a particle in Brownian motion, and then the random variation to be expected in radioactive decay. The arguments are set in the context of the general importance of randomness both in physics and in everyday life. We think that the ideas could usefully form part of students' A-level work on random decay and quantum phenomena, as well as being good for their general education. In the second article we offer a novel and simple approach to Poisson sequences.

Why random behaviour matters

We believe that random phenomena are unjustly neglected in A-level physics. Here are some reasons why they are important, in general education as well as in physics.

The newspapers report a 2% shift from 50% to 52% in support for some idea. Is this real news or not? Political polls usually sample about 2000 people. A good rule of thumb, and one of the ideas that this article is all about, says that random statistical fluctuations in N counts will often give rise to variations of $\pm\sqrt{N}$ counts. That's about ± 30 out of 1000 supporters, a change of 3%. A more careful analysis (see later) expects just over 2% variation. So maybe this news isn't news at all. Understanding things like this is one reason—the social reason—to believe that students ought to understand random behaviour better.

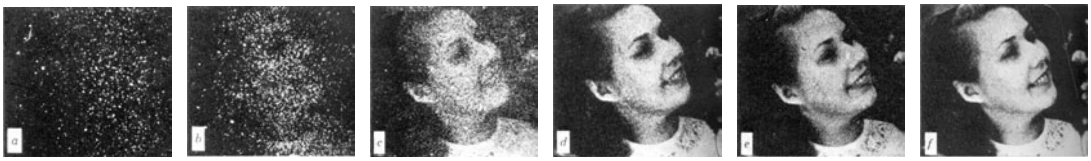
Nowadays, a photograph—such as that of a new supernova—is made by collecting photons in the pixels of a 'charge-coupled device' (CCD) like the one in your digital camera. Figure 1 shows how an image changes as the number of photons increases. With few photons the image is totally unclear, because the photons fall at random, with a probability depending on the intensity at each place. Just when can the astronomer feel sure that there really is a supernova, not just a chance fluctuation? This is another reason—the technological reason—why random behaviour matters.

Finally, random behaviour matters because it is at the root of quantum behaviour, which is to say, at the root of our deepest understanding of how things are. Quantum calculations predict, not what will happen, but the *probability* of events. This is the reason, for example, why radioactive decays come at random. It is also the reason why photons arrive at random. Yet in physics teaching, these principles are usually at best merely stated, not demonstrated. So a last and fundamental reason why understanding randomness matters is to be able to demonstrate its presence in real phenomena such as radioactive decay.

The drunkard's walk

Molecules in a gas dash hither and thither at random, changing direction unpredictably

0031-9120/03/050391+07$30.00



3×10^3 photons 1.2×10^4 photons 9.3×10^4 photons 7.6×10^5 photons 3.6×10^6 photons 2.8×10^7 photons

Figure 1. The same image, taken with increasing numbers of photons.

whenever they collide. Here is a prime example of random behaviour, with some surprising consequences. When Einstein showed how to understand it, it became possible for the first time actually to count molecules.

The simplest imaginable model of the motion is the simple 'drunkard's walk'. Imagine a drunk, taking steps of equal length in random directions. How far will he get from the starting point, and how will the distance vary? To keep things as simple as possible, we'll analyse the motion only in one dimension (the drunk staggers to and fro on a line—figure 2). This is enough to get the most important result: that the likely variation in a walk of N steps is equal to $\sqrt{N}$ steps.

Let the drunk take steps x_1, x_2, x_3 etc with each x equal to either +1 or −1, with equal probability. The total distance X gone after N steps is:

$$X = x_1 + x_2 + x_3 + \ldots + x_i + \ldots + x_N.$$

Each step x_i is equally likely to be +1 or −1. Thus on average, over many such walks, the total distance will add up to zero. This is called the *expected value of* X, written $\mathrm{E}(X)$. Thus:

$$\mathrm{E}(X) = 0.$$

Notice that the expected value can be written down directly from the model. Thus it is a theoretical value that can be compared with experiment. This is why we use the concept of expected values here.

In any given random walk, X can differ somewhat from the expected value, which is the value around which the average converges over many walks. By how much will X typically vary from the expected value? On any one occasion, the departure r from the expected value is just

$$r = X - \mathrm{E}(X).$$

It's no use asking for the expected value of r, because that is clearly zero[1]. But the 'spread' of the variations of X around the expected value can be looked at by considering the magnitude of r. The usual way to do this is to ask about the expected value of the (always positive) square of r. Since $\mathrm{E}(X) = 0$ this is:

$$\mathrm{E}(r^2) = \mathrm{E}(X^2) = \mathrm{E}(x_1{+}x_2{+}x_3{+}\ldots{+}x_i{+}\ldots{+}x_N)^2.$$

Think about multiplying out the squared bracket[2]. There will be N terms like x_i^2. There will also be 'mixed terms' of the form $x_i x_j$ with i and j different. There are $N(N-1)$ of them, because you can choose the first one in N ways and a second different one in $(N-1)$ ways. Note that they come in pairs, since $x_i x_j = x_j x_i$. The expected value of r^2 is the sum of expected values of these two kinds of terms:

$$\mathrm{E}(r^2) = \mathrm{E}\left(\sum_{i=1}^{N} x_i^2\right) + \mathrm{E}\left(\sum_{i \neq j} x_i x_j\right).$$

In the first sum, all the values of x^2 are equal to +1. There are N of them so the first sum adds up to N. In the second sum, the 'mixed products' can be $1 \times 1 = 1$, $-1 \times -1 = 1$, $1 \times -1 = -1$ or $-1 \times 1 = -1$. So they are equally likely to be positive as negative. Thus the expected value of their sum is simply zero. We get the very simple result

$$\mathrm{E}(r^2) = N.$$

The expected value of the squared departures r is called the *variance*. Its square root is the standard

steps +1 or –1 with equal probability

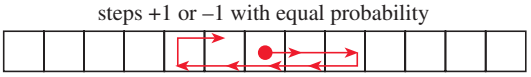


Figure 2. A random walk in one dimension.

[1] Formally, $\mathrm{E}(r) = \mathrm{E}[X - \mathrm{E}(X)] = \mathrm{E}(X) - \mathrm{E}[\mathrm{E}(X)] = \mathrm{E}(X) - \mathrm{E}(X) = 0$.

[2] Compare $(a+b)^2 = a^2 + ab + ba + b^2$.

deviation σ, a good measure of the spread[3]. Thus for the drunkard's walk:

$$\sigma = \sqrt{N}.$$

The meaning and use of $\sqrt{N}$

Molecules in a gas move in random directions, but in three dimensions. They go on average precisely nowhere (figure 3), but the root mean square of the distances they go in N collisions is just $\sqrt{N}$ steps. Typically such molecules make about 10^9 collisions each second, or 10^{12} collisions in a quarter of an hour. So molecules will quite often diffuse about $\sqrt{10^{12}} = 10^6$ steps in that time. The mean distance between collisions is typically 10^{-7} m, so you can expect a gas to diffuse about 0.1 m in that time. This contrasts with the enormous distance of 10^{-7} m $\times$ 10^{12} steps $=$ 10^5 m $=$ 100 km that a molecule travels along its tangled path.

Astronomers sometimes have to make do with very few photons arriving from a distant object in each pixel of the CCD attached to a telescope. How few can they manage with? Assume for now that the standard deviation is also equal to $\sqrt{N}$. Then if the average number of photons per pixel is 10, there will often be variations of 30%, and sometimes twice as much. There is no chance of a well-defined image. Even with 10 000 photons per pixel, the fluctuations still amount to 1%. Look back to figure 1 to see this illustrated.

A related fact is the reduction in experimental error through averaging many results. If there are N results with a spread σ, then the expected spread of the mean of those results is of the order $\sigma/\sqrt{N}$. The reason is simply the way random errors can as easily be negative as positive, just as in the previous argument. But notice that this strategy does not work at all for systematic error, which stays the same for each measurement.

The full theory of error and random variation is complicated and quite difficult. But you won't go far wrong if you hang on to the idea that N counts can easily vary by $\sqrt{N}$. By 'can easily vary' we mean that a majority (often around 2/3) of results fall within $\pm\sigma$ of the central value.

Notice especially that although the random variation $\sqrt{N}$ increases as N increases, its relative value $\sqrt{N}/N = 1/\sqrt{N}$ *decreases*. More counts means greater precision. But that extra precision may be too costly to obtain, because the gain in precision increases much more slowly than the increase in the number N of values required to achieve it.

Random radioactive decays

Because radioactive decay is a quantum phenomenon, counts from radioactive decay arrive at random. In the case of a material with a long half-life, the probability p of arrival in a given short time Δt is constant. Writing λ for the constant, the probability per unit time, you get

$$p = \lambda \Delta t.$$

Again we ask: what is the mean number of counts in time t, and by how much can the number of counts be expected to vary? The answers can be approached through another random walk argument.

Random walk: the timid traveller

We invented the following argument, having found nothing simpler elsewhere. But it seems very unlikely that we are actually the first to have thought of it.

Picture a nervous traveller. At successive moments of time, the traveller either takes one

[3] There are other ways of measuring 'spread'. However, as N varies, all of them scale as $\sqrt{N}$, which is the key result here.

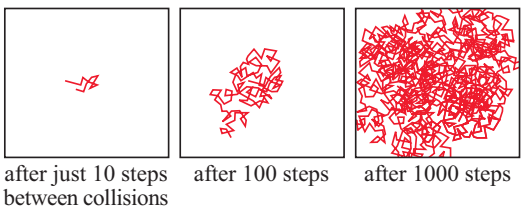


Figure 3. Simulated Brownian motion: the path of a molecule.

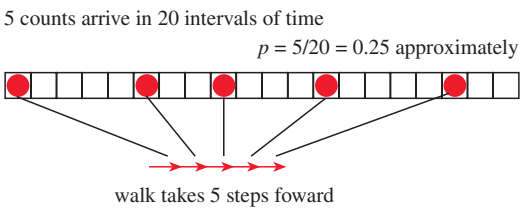


Figure 4. Random arrival of counts modelled by a random walk.

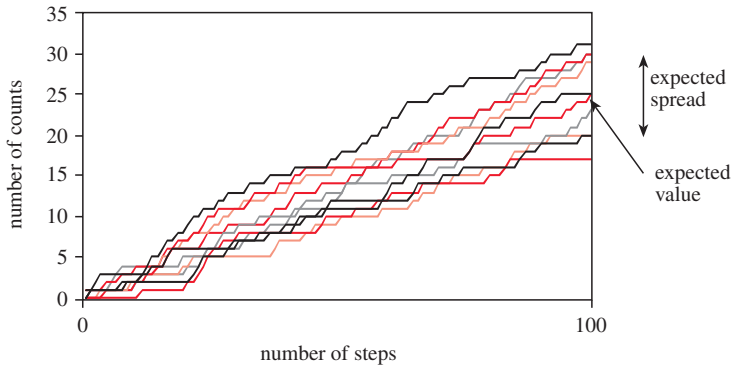


Figure 5. Ten sample random walks ($p = 0.25$, $N = 100$): expected mean 25; expected standard deviation 5.

step forward or stays irresolutely still. Let the probability to take a step forward be p. Then the probability to stay still is $1 - p$. In this picture, taking a step forward corresponds to one count arriving. Staying still corresponds to no count arriving (figure 4).

How far will the timid traveller get on average, and by how much will the distance gone vary from one occasion to another? We will suppose that the traveller makes a random decision N times, at time intervals Δt, and that the steps taken have unit length. The difference between this situation and radioactive decay is that we have chopped up time into discrete intervals, in each of which one event can happen or not.

Average path in chopped-up time

Let x_i be the length of the ith step. Then $x_i = 1$ with probability p and $x_i = 0$ with probability $1 - p$. The total length of the walk is given by

$$X = x_1 + x_2 + x_3 + \ldots + x_i + \ldots + x_N.$$

What is its expected value over many such walks? On average, there will be Np steps equal to 1, and $N(1 - p)$ steps equal to 0. Thus the expected length of a walk (or total number of counts) is[4]

$$\mathrm{E}(X) = Np.$$

Thus, in 100 possible steps, if the probability of actually taking a step is 0.1, the expected number of steps actually taken is 10.

[4] To spell it out: $\mathrm{E}(X) = 1 \times Np + 0 \times N(1 - p) = Np$.

Variation of length of path in chopped-up time

Figure 5 shows a simulation of the timid traveller's walk. You can see how the total distance gone varies around an average value. We want to know how much the length of the walk (total number of counts) is likely to vary. The difference r between the length of any walk and the expected length is

$$r = X - \mathrm{E}(X).$$

The difference r is often called the *residual*. Clearly r is just as likely to be negative as positive, so its expected value is zero. But the square of r is always positive, and we can ask (as previously) about the expected value of its square:

$$r^2 = (X - \mathrm{E}(X))^2.$$

You get a very generally useful algebraic fact if you multiply out the squared bracket and ask about the expected value of each term:

$$r^2 = X^2 - 2X\mathrm{E}(X) + (\mathrm{E}(X))^2.$$

The middle term is just X multiplied by twice the expected value of X. So its expected value is $-2(\mathrm{E}(X))^2$ and the expected value of r^2 comes to[5]

$$\mathrm{E}(r^2) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$$

The expected value of squares of residuals is the variance. Mean values over a long run are good estimates of expected values. So remember the tongue-twisting mnemonic for calculating an experimental estimate of the variance:

[5] $\mathrm{E}(r^2) = \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + (\mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

"the mean of squares of X minus the square of the mean of X."

You already know the square of the expected value of X: it is $(Np)^2$. So to find the variance you just need to find the expected value of X^2 and subtract $(Np)^2$ from it.

The square of the path length (number of counts) X is

$$X^2 = (x_1 + x_2 + x_3 + \ldots + x_i + \ldots + x_N)^2.$$

As before, squaring the long bracketed expression gives two kinds of term: N terms of the form x_i^2 and $N(N-1)$ 'mixed' terms of the form $x_i x_j$ where i and j are different. That is

$$X^2 = \sum_{i=1}^{N} x_i^2 + \sum_{i \neq j} x_i x_j.$$

The expected value of X^2 is the sum of the expected values of the two sets of terms. In the first set, on average there will be Np cases where $x_i = 1$ and of course also $x_i^2 = 1$, so the expected value of the first sum of terms $= Np$.

The 'mixed' terms can have the form 1×1, 1×0, 0×1 or 0×0. Only the first kind have value 1, the rest being zero. How many of the first kind are there? x_i takes the value 1 with probability p, so two of them independently both have the value 1 with probability p^2. Thus out of $N(N-1)$ terms, a total of $p^2 N(N-1)$ will on average have the value 1. Therefore the expected value of the second sum is $p^2 N(N-1)$.

This gives the expected value of the squares of X:

$$\mathrm{E}(X^2) = Np + N(N-1)p^2$$

or, rearranging,

$$\mathrm{E}(X^2) = Np(1-p) + (Np)^2.$$

To find the variance, just subtract $(Np)^2$, the square of the expected value of X, giving

$$\mathrm{E}(r^2) = Np(1-p).$$

The standard deviation σ is therefore

$$\sigma = \sqrt{Np(1-p)}.$$

This is an exact and very valuable general result, with many uses wherever there are N random choices between two possibilities, that is, for any binomial distribution. In the case discussed above, with $N = 100$ and $p = 0.1$, the standard deviation is

$$\sigma = \sqrt{100 \times 0.1 \times (1 - 0.1)} = \sqrt{9} = 3.$$

Whilst the expected length of the walk is 10 units, the actual length in any one case can easily vary in the range 10 ± 3 units.

Another example: if you toss a coin N times, with probability $p = 0.5$ of a head, then the average number of heads $N_{\mathrm{average}} = Np = 0.5N$. The standard deviation of the number of heads is

$$\sigma = \sqrt{N \times 0.5 \times 0.5} = \sqrt{0.5 N_{\mathrm{average}}}.$$

So in 100 tosses you expect 50 heads, but this can easily vary by $\pm\sqrt{25} = \pm 5$. A similar calculation estimates the expected variation in the results of opinion polls (see first page).

The result is very important in digital data collection (e.g. imaging) in which a sensor either receives a pulse or not. It determines the random variation to be expected in the response of the sensor. Another use is in tracking failure rates of components on microchips. In the fabrication process, some components suffer random defects. It is important to know how many of these to expect, and also to know by how much the number of defects may vary. Exact results that are useful enough to be worth remembering are rare in physics. This is one of them.

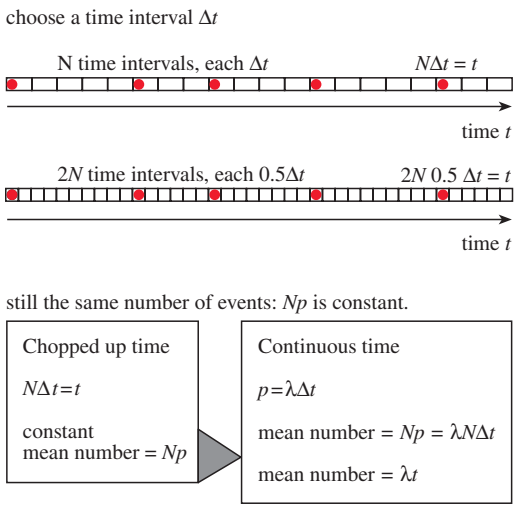


Figure 6. From chopped-up time to continuous time.

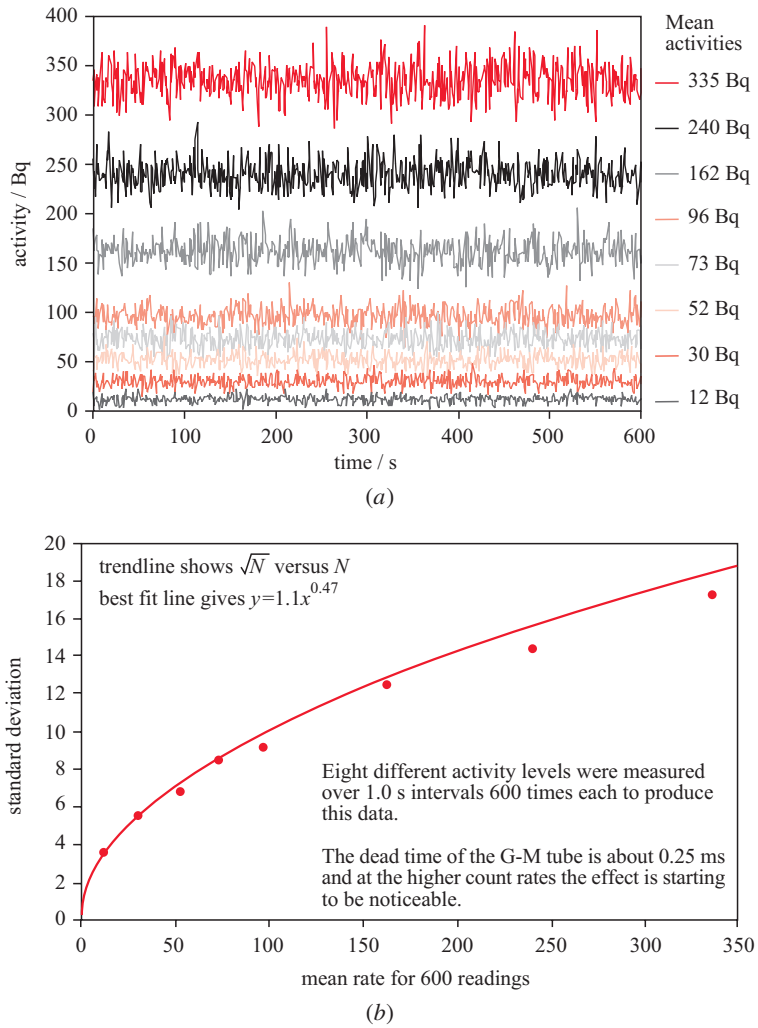


Figure 7. (*a*) Different levels of activity, with different spreads. (*b*) The standard deviation of N counts varies as $\sqrt{N}$ (data as in (*a*)).

Going back to continuous time

In radioactive decay, or in the random arrival of photons, there is really no such thing as the 'number N of intervals Δt in which there may be a decay'. The previous picture chopped a total time t into N imaginary slices each of length Δt, such that $N\Delta t = t$.

In reality time runs continuously, with the probability p of a count in a short time Δt being given by

$$p = \lambda \Delta t.$$

As you imagine the time slices Δt getting smaller and smaller (see figure 6), the probability p gets smaller and smaller. Thus $(1 - p)$ becomes closer and closer to 1. In this case, the standard deviation can be written simply as

$$\sigma = \sqrt{Np}.$$

Now you can use the expression for p in terms of the rate λ to write this in a better way, getting rid of the imaginary number N of time slices altogether

$$Np = \lambda N \Delta t = \lambda t$$

so that

$$\sigma = \sqrt{\lambda t}.$$

Remembering that $Np = \lambda t$ is also equal to the expected number of counts, this can also be

remembered as

$$\sigma = \sqrt{N_{\text{expected}}}.$$

These last results are also exact, for the continuous random process of arrival of counts in time. Thus a Geiger counter showing an average count rate of 50 Bq from a source will give an expected 1000 counts in 20 s, but this will vary over at least the range $1000 \pm \sqrt{1000} = 1000 \pm 30$ approximately. Other examples include the rate of arrival of telephone calls at an exchange, the rate of occurrence of errors in a message, and the rate of arrival of photons from a star, at the detector in a telescope. An up-to-date example is the expected rate of occurrence of detections of neutrinos from a supernova, and the variation to be expected in that rate.

Experimental check on $\sqrt{N}$

It is not difficult to use a counter and fast data-logger to collect samples of numbers of counts in a fixed time. Figure 7(a) shows sample data taken in this way. Counts were repeatedly taken for a fixed time, at a variety of different levels of activity. For each activity level, the mean number of counts N was recorded, and the actual standard deviation of the variation between the samples at that activity was calculated. You see from figure 7(b) how well the prediction that the standard deviation is equal to $\sqrt{N}$ works out experimentally. Notice how at high rates the 'dead time' of the counter begins to have an effect.

These data give some experimental reason to believe that the counts from a radioactive source—a true quantum phenomenon—do really arrive randomly. And other valuable lessons about random variations and experimental error are clear to see.

The graph of figure 7(b) shows clearly how $\sqrt{N}$ increases less rapidly than N. Thus although the variation *increases* as N increases, the proportional variation *decreases*, so the variation becomes less and less important.

There's an important message here for estimating how many counts you need to get an accurate estimate of the mean rate of arrival. For example, at the highest activities of about 335 Bq, the expected spread on a single one-second reading is about $\sqrt{335} \simeq 18$, about 5%. In fact 600 repeated readings are taken, so that the expected spread in the mean is $\sqrt{335 \times 600} \simeq 448$, which is 0.2%. A 600-fold increase in effort only yields a 25-fold improvement in precision. $\sqrt{N}$ strikes again! Similar ideas are important for astronomers counting photons in pixels of an image.

Conclusion

We think that the basic ideas here, using the idea of a random walk and getting the value $\sqrt{N}$ for the standard deviation, are accessible to most A-level students, and are worth attempting on account of their general importance.

This is as far as many A-level teachers of physics will want to go. But it is worth thinking about going a bit further. Doing so delivers some surprising results and suggests some further simple and interesting experiments, looking for example at how the waiting times between random counts vary—a question with applications to, for example, queuing at the supermarket.

The following article takes up this theme, developing the concept of Poisson sequences in a novel and simple way, and offering further tests of the randomness of a process. The goal is a better insight into how random events work.

Received 13 May 2003, in final form 11 June 2003
PII: S0031-9120(03)63337-X

Jon Ogborn directed the Institute of Physics' *Advancing Physics* project, and is Emeritus Professor of Science Education, Institute of Education, University of London. **Mick Brown** is a Fellow of the Royal Society, Emeritus Professor of Physics, University of Cambridge, and has a long-standing interest in improving physics education. **Simon Collins** teaches physics at Stowe School, and is a Principal Examiner for *Advancing Physics*.