Feynman Lectures, Chapter 6: Probability

David Mummy

December 18, 2017

We first define probability of an outcome (6-1) as *"our estimate for the mostly likely fraction of a number of repeated observations that will yield that particular outcome."* In mathematical terms, if we imagine tossing a coin N times, and if we estimate based on our prior knowledge that outcome A will occur N_A times, then we say that $P(A)$, the probability of A occuring, is

$$P(A) = N_A/N$$

For an fair coin $P(A)$ will be $\frac{1}{2}$ whether A is heads or tails, since we expect them to occur the same number of times (namely $N_A = N/2$). By this same logic, if there are m equally-likely outcomes, the probability of any of the individual outcomes is $\frac{1}{m}$. This formulation of course only makes sense for something where we can make repeated observations – or at least, multiple observations that we consider *equivalent*. (We assign the probability of rain on a given day, but obviously no two days are exactly alike).

Of course, if we flip a coin 30 times and get 17 heads, it doesn't mean we are wrong; 15 heads the most likely outcome out of all the outcomes of 30 coin flips, but the results will fluctuate (6-2). We can perform repeated observations and sketch out the *probability distribution* – the relative number of results for each outcome. How do we figure out these probabilities? We can use a sort of modified decision tree (or an inverted March Madness bracket if you prefer) to map out the different outcomes, where outcomes now are *combinations* of tosses, not a single toss. We start with the first toss, and it branches two ways (heads or tails), each of those branches then branches, reflecting the second toss, and so on. So for a sequence of three tosses, there is only 1 path to three heads (H-H-H) but three paths to 2 heads (H-H-T, H-T-H, T-H-H), and similarily 3 for 2 tails, and 1 for all tails. So there are $1+3+3+1 = 8 = 2^3$ total paths, with some paths yielding the same result. Thus the probability of getting all heads is 1/8, but there are three ways/paths to get exactly 2 heads, so the probability of exactly 2 heads is 3/8. Note that since each branch yields 2 more branches, we get a power of 2 number of outcomes.

1-3-3-1 may ring a bell; we are just recreating Pascal's triangle here. These are also called the *binomial coefficients* since they are the coefficients in the expansion of $(a+b)^n$. That's not a coincidence; we're also counting combinations when we multiply out the binomial. We can write the coefficients using the "n

choose k" notation thus:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

.

We can generalize our coin-game, writing that the probability of getting k heads in n tosses is

$$P(k,n) = \frac{\binom{n}{k}}{2^n}$$

We can further generalize this to a game where the coin is "rigged", i.e. each branching is not equally likely. If, at each flip, one outcome is probability p and the other is q (which is by necessity $1-p$), we can write

$$P(k,n) = \binom{n}{k} p^k q^{n-k}$$

Think of the p's and q's as reflecting the accumulating probabilities as we move through a given path on the tree; the $\binom{n}{k}$ represents the number of paths leading to that same outcome. Note that if p and q are both $1/2$, this reduces to the previous equation.

We abstract this further into the idea of a *random walk* (6-3): a player starts at 0 and then makes a equally likely move forward or backward one "unit". How far away does she get after N moves? We don't care about negative and positive, so we consider the square of her distance D^2. Here we have a nice proof of the average "drift" from the starting spot. It is clear that D_1^2, the distance after the first move, is 1. And we know after N moves, $D_N = D_{N-1} - 1$ or $D_N = D_{N-1} + 1$. If we multiply this out, we get

$$D_N^2 = \begin{cases} D_{N-1}^2 + 2D_{N-1} + 1, \\ \qquad or \\ D_{N-1}^2 - 2D_{N-1} + 1, \end{cases} \tag{1}$$

Both outcomes are equally likely, so our expectation is the expacted value of the average of the two values: $\langle D_N^2 \rangle = \langle D_{N-1}^2 \rangle + 1$. We know $\langle D_1^2 \rangle = 1$, so if we just iterate that forwards, we get $\langle D_N^2 \rangle = N$. So that's cool. If we want to see an actual distance, we can use the root mean square:

$$D_{rms} = \sqrt{\langle D^2 \rangle} = \sqrt{N}$$

Some algebra shows that the number of heads N_H goes as

$$N_H - \frac{N}{2} = \frac{D}{2}$$

and thus the expected rms difference between N_H and its expected value $\frac{N}{2}$ is $\left(N_H - \frac{N}{2}\right)_{rms} = \frac{1}{2}\sqrt{N}$. We thus expect the ratio of the deviation to the total N $\frac{\sqrt{N}}{2})\frac{1}{n} = \frac{1}{2\sqrt{N}}$ to approach 0 as N gets large – in other words, the number

of heads approaches 0.5. We can incorporate this into an "expectation" even for an unfair coin by writing the probability of heads $P(H)$ incorporating a "fluctuation term" as

$$P(H) = \frac{N_H}{N} \pm \frac{1}{2\sqrt{N}}$$

This of course depends on the coin being at least somewhat fair (i.e. $P(H)$ is near $\frac{1}{2}$) since that's where the 2 came from in the first place.

What if we let the step size S vary but keep $\langle S^2 \rangle = 1$? We still get $\langle D_N^2 \rangle = N$. Now, though, the likelihood of any *particular* displacement is zero (there are infinite "paths"), so we turn to the idea of a probability density (6-7) $p(x)$

$$P(x, \Delta x) = p(x)\Delta x.$$

and we can integrate it over some interval (x_1, x_2) to sum the infinite "paths" and determine the likelihood of the deviation occuring over that particular interval. The integral of $p(x)$ over $(-\inf, \inf)$ must of course be 1.

This particular probability density is a *normal* or *gaussian* probability density:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -x^2/2\sigma^2$$

with standard deviation σ; for our case $\sigma = \sqrt{N}$ but more generally $\sigma = \sqrt{N}S_{rms}$.

These probability densities (and the integrations thereof) can be used to describe distributions of velocities of gas particles (more on this later). You can think of these delta intervals of probability densities as the δx vs. δp etc. presented in the uncertainty equations we saw in previous chapter. When we speak about the components of atoms, we must speak in probability densities.