Physics 127a: Class Notes

Lecture 2: A Simple Probability Example

The "equally likely" of the fundamental postulate reminds us of a coin flip, and in fact a very simple probability problem actually gives us useful insights into statistical mechanics issues. The problem is: What is the probability of getting m heads in a sequence of N flipped coins? We will denote this $P(m, N)$. In particular, we will compare the probability of getting $N/2$ heads ($m = N/2$) with the probability of getting N heads ($m = N$).

This is actually the same problem as a number of simple, but not uninteresting, statistical mechanics problems:

- For a magnetic system of N noninteracting magnetic moments that can each point either "up" or "down" (e.g. a set of spin-$\frac{1}{2}$ atoms) what is the probability of a state with m up moments and $N - m$ down moments, i.e. a magnetization of $2m - N$ times the individual moment?
- Given an ideal gas of N molecules, what is the probability of finding m molecules in one half of the box, and $N - m$ molecules in the other half, and in particular the probability of an equal number of molecules in each half, compared with the probability of all the molecules being in one half?
- A drunkard's walk along a path, or a "one-dimensional random walk". At each step the drunkard may go one step forwards or one step backwards, with equal probability. What is the probability of finding the drunkard at position x after N steps? This is $P(m, N)$ with $x = 2m - N$ (m forward steps and $N - m$ backward steps).

Back to the coin problem. Consider first $N = 4$ coins. Since any sequence, e.g. HHTT or HTHT, is equally likely (assuming unbiased coins) we can calculate $P(m, 4)$ by counting the number of sequences or "microstates" that are consistent with each "macrostate" m

m	microstates	no.	$P(m, 4)$
0	TTTT	1	$\frac{1}{16} = 0.0625$
1	HTTT,THTT,TTHT,TTTH	4	$\frac{4}{16} = 0.25$
2	HHTT,HTHT,THHT HTTH,THTH,TTHH	6	$\frac{6}{16} = 0.375$
3	THHH,HTHH,HHTH,HHHT	4	$\frac{4}{16} = 0.25$
4	HHHH	1	$\frac{1}{16} = 0.0625$

Already $P(N/2)$ is several times $P(N)$, and the ratio increases rapidly as N increases.

The general expression for $P(m, N)$ for two outcomes A and B with individual probabilities p_A and $p_B = 1 - p_A$ is

$$P(m, N) = p_A^m p_B^{N-m} \frac{N!}{m!(N-m)!} \tag{1}$$

known as the *binomial distribution*. Here the first factor is the probability of a particular sequence with m outcomes A, and the second factor counts how many such sequences there are. The coin problem is this result with $p_A = p_B = \frac{1}{2}$. It is then possible to show directly that for large N (and see below) $P(N/2, N) = \sqrt{\frac{2}{\pi N}}$ and $P(N, N) = \frac{1}{2^N}$ so that for large N we see $P(N/2, N) \ggg P(N, N)$.

You will investigate $P(m, N)$ for large N using *Stirling's approximation* for factorials of large numbers in the homework. We can actually get the interesting properties from a couple of simple arguments and a profound result. The simple arguments are:

Mean: The mean or average value of the number of heads is

$$\langle m \rangle = \left\langle \sum_{i=1}^{N} x_i \right\rangle \tag{2}$$

where x_i is a random variable which takes on the values 1 (heads) with probability $\frac{1}{2}$ and 0 (tails) with probability $\frac{1}{2}$. The $\langle \, \rangle$ stand for the ensemble average. We can interchange the order of the sum and the average, so

$$\langle m \rangle = \sum_{i=1}^{N} \langle x_i \rangle = \frac{N}{2}. \tag{3}$$

Variance: The variance or mean square fluctuation in the number of heads is

$$\sigma_N^2 = \left\langle \left(\sum_{i=1}^{N} x_i - \langle m \rangle \right)^2 \right\rangle = \left\langle \left(\sum_{i=1}^{N} \left(x_i - \frac{1}{2} \right) \right)^2 \right\rangle \tag{4}$$

$$= \left\langle \sum_{i=1}^{N} \left(x_i - \frac{1}{2} \right) \sum_{j=1}^{N} \left(x_j - \frac{1}{2} \right) \right\rangle \tag{5}$$

$$= \left\langle \sum_{i=1}^{N} \left(x_i - \frac{1}{2} \right)^2 + \sum_{i=1}^{N} \sum_{\substack{j=1,\\ j \neq i}}^{N} \left(x_i - \frac{1}{2} \right) \left(x_j - \frac{1}{2} \right) \right\rangle. \tag{6}$$

Interchanging the order of averaging and summing as before, the first term is $N/4$ (N terms each equal to $\frac{1}{4}$) and the second gives zero since the two factors are both equally likely to be $\pm\frac{1}{2}$. So $\sigma_N = \sqrt{N}/2$, which gives the "width" of the distribution.

Finally, we use the **central limit theorem**, which tells us that the probability distribution of a quantity Q formed as the sum of a large number of random variables with *any* distribution (with finite mean and variance) is *Gaussian*, i.e.

$$P(Q) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Q-\bar{Q})^2}{2\sigma^2}} \tag{7}$$

with $\bar{Q}$ the mean and σ^2 the variance.

Thus for large N for the coin toss problem we have, since m is the sum of N random variables x_i,

$$P(m, N) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\frac{(m-N/2)^2}{2\sigma_N^2}} \tag{8}$$

with $\sigma_N = \sqrt{N}/2$ as we have just calculated. Note that this expression for $P(m, N)$ is accurate only for m not *too* far from the most probable value $N/2$ (i.e. not too many σ_N away) — which however is the only region where the probability is significantly nonzero. Where the probability is *very* small the result is inaccurate. For example for $m = N$ we know $P(N, N) = 2^{-N} = e^{-N \ln 2}$, which is very different from the result given by Eq. (8). As N gets large, the width of the probability distribution of the number of heads m also becomes large, proportional to $\sqrt{N}$. But this is *small* compared to the range N, and becomes *very small* in this comparison for N equal to the number of molecules in a macroscopic object, say 10^{24} (cf. the gas-in-the-box problem).

The the number of heads is an *extensive* quantity — the mean value grows proportional to the size of the system N. It is often convenient to introduce an intensive variable, such as the fraction of heads $f = m/N$, with a mean that does not grow with system size. Since it is natural to consider f as a continuous variable for large N, we introduce the *probability density* $p(f)$ such that the probability of a fraction between f and $f + df$ for small df is $p(f)df$. Then since $P(fN, N) = p(f)\frac{1}{N}$ we have

$$p(f) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(f-1/2)^2}{2\sigma^2}} \tag{9}$$

with $\sigma = \frac{1}{2\sqrt{N}}$. For the intensive variable the probability distribution gets narrower proportional to $N^{-1/2}$ as N gets large. Again the Gaussian distribution is only good not too many σ away from the most probable value, i.e. for f within of order $N^{-1/2}$ of $1/2$ — but again this is the only region where $p(f)$ is significantly nonzero. In fact the probability of any fraction f not equal to one half (e.g. $f = 1/3$ or $f = 1/29$) is *exponentially small* for large N, i.e. of order e^{-aN} with a a number of order unity that depends on f but not N. For example the probability of finding on one side of the box a fraction f of gas molecules that is not equal to one half (e.g. one third) is of order

$$P(f \neq \frac{1}{2}) \sim 10^{-10^{24}} \tag{10}$$

i.e. $10^{-1000000000000000000000000}$. (Note $a \times 10^{24}$ is of order 10^{24} for any reasonable number a!) This is a number that for *all* physical purposes is zero (much, much, much smaller than one over the number of atoms in the universe etc.).

This way that probabilities become certainties for N large corresponding to the number of molecules in a macroscopic sample, will be a recurring theme in the application of statistical mechanics of macroscopic systems. In general we will find for the probability distribution of macroscopic quantities:

- For extensive quantities with a mean proportional to N (e.g. the total energy) fluctuations that are relatively small of order $\sqrt{N}$ with a Gaussian distribution about the mean;
- For intensive quantities with a mean of order unity (e.g. the temperature) small fluctuations of order $1/\sqrt{N}$ again with a Gaussian distribution about the mean
- The distribution is so narrow that we can replaces averages of quantities over the distribution by the result evaluated at the most probably value.
- Fluctuations far away from the mean have a probability that is exponentially small in N, and for $N \sim 10^{24}$ can be considered as *never* happening.