1. Introduction

1.1 What is Statistical Mechanics?

Thermal Physics encompasses all parts of physics where ideas of temperature and entropy come into play. As we shall see this implies the properties of macroscopic systems with a large number of microscopic constituents.

Macroscopic Approach (Classical Thermodynamics)

- deals with macroscopic variables i.e. variables that do not refer to any microscopic details
- input is phenomenological laws e.g. equation of state
- output is general relations between macroscopic variables
- advantage is the generality of the approach

Microscopic Approach (Statistical Mechanics)

- starts from a microscopic description and seeks to explain macroscopic properties.
- input is a microscopic model of a given system i.e. what we believe is an adequate microscopic description
- output is predictions for macroscopic properties and behaviour.
- advantage of the approach is that it yields predictions for a given system which can be compared to experiment thus allowing refinement of the microscopic model which in turn deepens our understanding of the system.

> The key idea that connects the microscopic and macroscopic approaches is that due to the large number of consituents, macroscopic quantities are precisely defined even though the microscopic specification is not precise.

Aims of course

- To formulate the ideas of statistical mechanics
- To give microscopic understanding of entropy & second law of thermodynamics which you met in Year 1 Properties of Matter (PoM)
- To derive results you have met in PoM such as equipartition and the Boltzmann factor
- To explore the role of indistinguishability and quantum effects in many body systems

1.2 Microscopic approaches

Extreme philosophy: If we know microscopic laws e.g. Newtonian mechanics; quantum mechanics etc then we just specify the initial conditions and solve all the equations to determine what happens!

Problems with this approach:

- Analytically intractible
- even on a powerful computer one can only deal with a limited number of consituents say $N \sim 10^5$, whereas Avogadro's number is $N_A = \sim 6 \times 10^{23}$.
- system may not even be deterministic! i.e. small change in initial conditions leads to completely different final states

Instead we will use the simpler and more powerful *statistical* formulation

- Only aim to make probabilistic statements e.g. the Maxwell probability distribution for the velocities of molecules in an ideal gas (see PoM) doesn't tell us about trajectories of individual gas molecules. However it does give us all we need to know to deduce macroscopic properties such as the pressure of the gas.

- It is a very natural and economical approach i.e. we only introduce as much detail as we need (or want) in the microscopic description and we do not waste time calculating unnecessary details. Therefore we can do the calculations analytically (i.e. on pen and paper).

- It uses the large number N of microscopic constituents to advantage

- Why the name 'statistical mechanics'?

  - The microscopic model is the 'mechanics' e.g. quantum, Newtonian or even more primitive (but adequate) descriptions
  - 'statistical' refers to the statistical description of the resulting behaviour of a system with a large number of microscopic constituents

Statistical mechanics was founded in the late nineteenth century with the work of Maxwell, Boltzmann, Gibbs ...on classical systems. It was developed hand in hand with quantum mechanics in the twentieth century by Einstein, Bose, Dirac, Fermi . . . It remains an active and constantly developing research area in the 21st century. For example the techniques of statistical mechanics are applied to neural modelling, the study of traffic flow, economic and social systems... in fact any system with a large number of constituents.

More detailed comparison of the two approaches (taking example of a gas):

	Thermodynamics	Statistical Mechanics
Quantity of interest	Macroscopic properties (eg. P, V, T, C_P, C_V)	Microscopic properties (eg: molecular speeds)
Strategy	Avoid microscopic model	Build on microscopic model
Strengths	Generality of results	Provides way of refining microscopic understanding
Weaknesses	No understanding of system-specific features. Conceptually opaque.	Requires additional input (the model). Requires additional techniques (probability theory; classical and quantum mechanics)

1.3 Probability

Probabilitic concepts are central to statistical mechanics. You should have a working knowledge of probability theory from your programme so far. We shall not explicitly revise this material in lectures, so refer to you probability notes as needed. For convenience and reference we summarise below some of the main results on which we shall draw.

Expand to review key concepts of probability

For further reading see chapter 3 of Gould and Tobochnik
Definitions

1. frequency definition: the probability P of an event in a trial is given by

$$P = \lim_{N\to\infty} \frac{n}{N}$$

   where n is the number of occurrences in N trials.

2. degree of belief: probability is a quantitative measure of our degree of belief that something will occur e.g. if there are q possible outcomes of a trial (recall that a trial is some procedure where we can measure the outcome), and we have no reason to favour any one outcome over any other, then we would assign probability $1/q$ to each outcome. Tossing a coin would correspond to $q = 2$ and rolling a die to $q = 6$.

Thankfully both definitions lead to the same numerical values for probability.
Rules of Probability
For a trial with q possible outcomes let the probability of outcome i be $P(i)$

- normalisation $\sum_{i=1}^{q} P(i) = 1$

- $P(i \text{ or } j) = P(i) + P(j)$ (for mutually exclusive outcomes i, j).

- We can also consider 'compound events'. For example consider two trials then

$$P(i \text{ in trial 1 and } j \text{ in trial 2}) = P(i)P(j)$$

(for outcomes of trial 1 and trial 2 independent)

Probability distributions

If we associate a numerical value x_i to event i then the probability distribution for the 'random variable' x is

$$P(i) = \text{Probability that } x = x_i$$

Probability density functions

In many case the outcome of an event is described by a continuous variable x. Then the probability density function $P(x)$ is defined by

$$P(x)dx = \text{Probability that outcome lies in the range } x \text{ to } x + dx$$

and is normalised according to

$$\int_{-\infty}^{\infty} P(x)dx = 1$$

(N.B. often the probability density function is simply called the probability distribution.)

Averages

The Probability distribution contains the complete information about the trial. However the moments (mean, variance . . .) can give the important features. They are defined as

$$\begin{aligned}
\text{mean } \ \overline{x} \quad &= \quad \sum_{i=1}^{q} P(i)x_i \ \text{ or } \int_{-\infty}^{\infty} P(x)xdx \\
\text{variance } \ \overline{\Delta x^2} \quad &= \quad \sum_{i=1}^{q} P(i)(x_i - \overline{x})^2 \ \text{ or } \int_{-\infty}^{\infty} P(x)(x - \overline{x})^2 \\
i.e. \ \overline{\Delta x^2} \quad &= \quad \overline{(x - \overline{x})^2} = \overline{x^2} - \overline{x}^2
\end{aligned}$$

More generally the average of a function $f(x)$ is given by

$$\overline{f(x)} = \sum_{i=1}^{q} P(i)f(x_i) \quad \text{or} \quad \int_{-\infty}^{\infty} P(x)f(x)dx$$

Basic distributions

For completeness (and for the mathematically inclined) the content below includes some derivations of the functional forms of distributions; you are not expected to know these derivations for this course.

The binomial distribution

First let us review the concepts of permutations and combinations. If we have N distinguishable objects—say numbered balls, or a pack of cards—then if we select n objects from N (without replacement) how many different possible outcomes are there?

The answer is known as the number of permutations of n from N and is equal to

$$\frac{N!}{(N-n)!} = \frac{N(N-1)(N-2)\dots(1)}{(N-n)(N-n-1)\dots(1)} = N(N-1)(N-2)...(N-n+1).$$

To see this think of the number of possibilities for the first object ($= N$) then multiply by the number of possibilities for the second $(N-1)$ and keep going until the nth object for which there are $(N-n+1)$ possibilities.

Also recall that $0! = 1$, so that the number of permuations of N from N is $N!$

In this counting scheme we in fact count as different outcomes the selection of the same objects but in different order e.g. clearly in the number of permutations of N from N we always choose the same N objects but the number of possible orders is $N!$

Now if we only count as distinct outcomes those events where a different set of objects is selected we are interested in the number of combinations of n from N. To obtain the number of combinations we have to divide the number of permutations of n from N by $n!$ (the number of permutations of the n selected objects). This results in

$$\text{number of combinations of } n \text{ from } N = \frac{N!}{(N-n)!n!} = \binom{N}{n}$$

where $\binom{N}{n}$ is known as a binomial coefficient and is pronounced 'N choose n'. (N.B. There are many other symbols that are used for the number of combinations.)

Now we are ready to write down the binomial distribution. We consider N trials in each of which an event can occur with probability p. The probability of observing precisely n events in N trials is given by the Binomial distribution

$$\boxed{P_n = \binom{N}{n} p^n (1-p)^{N-n}}$$

The binomial coefficient gives the number of ways of choosing the n trials where an event occurs out of the total of N trials.

The factor $p^n(1-p)^{N-n}$ is the probability that in the n chosen trials an event occurs and in the rest of the trials an event does not occur. (This factor is the same whatever the n chosen trials). This argument may seem somewhat obscure at first but should with familiarity become second nature.

Note that for the Binomial distribution

$$\boxed{\overline{n} = Np \quad \text{and} \quad \overline{\Delta n^2} = Np(1-p)}$$

It should be pointed out that the factorials involved in combinatorial calculations soon become very large numberse.g. for $N = 0, 1, 2 \ldots$ one can check that $N!$ is $1, 1, 2, 6, 24, 120, 720 \ldots$ and already $15! \sim 10^{12}$.

A very useful approximation for the factorial function is given by the *Stirling approximation*

$$\ln(N!) = N \ln N - N + \frac{1}{2} \ln(2\pi N) + O(1/N)$$

We see that for N large

$$\boxed{\ln(N!) \simeq N \ln N - N}$$

Poisson distribution

As an application of Stirling's approximation let us consider the binomial probabilities for $N \gg n$. Then we find

$$\begin{aligned}
\ln \binom{N}{n} &= \ln N! - \ln(N-n)! - \ln n! \\
&\simeq N \ln N - N - (N-n)\ln(N-n) + (N-n) - \ln n! \\
&\simeq n \ln N - \ln n!
\end{aligned}$$

In the calculation we used Stirling's approximation for $\ln N!$ and $\ln(N-n)!$ and made use of the fact that for $N \gg n$, $\ln(N-n) \simeq \ln N - n/N$. Note that we have only kept the highest order terms. Thus, re-exponentiating the logarithm, we see that

$$\binom{N}{n} \simeq \frac{N^n}{n!}$$

Now consider

$$(1-p)^{N-n} = \exp[(N-n)\ln(1-p)] \simeq \exp(-Np)$$

where we have used $\ln(1-p) \simeq -p$ for $p \ll 1$ and $N - n \simeq N$. (More strictly we require $N \to \infty, p \to 0$ such that $Np = \overline{n}$ is finite.) Under these conditions we find when we put the above results together that

$$\boxed{P_n = (Np)^n \frac{\exp(-Np)}{n!} = (\overline{n}^n) \frac{\exp(-\overline{n})}{n!}}$$

This is known as the Poisson distribution and one can check using the power series representation of the exponential function that $\sum_{n=1}^{N} P_n = 1$. One also finds

$$\boxed{\begin{aligned}
\overline{n} &= \sum_{n=1}^{N} n P_n = Np \\
\overline{\Delta n^2} &= \sum_{n=1}^{N} (n - \overline{n})^2 P_n = Np
\end{aligned}}$$

ie. the mean and variance are equal.
Basically the Poisson distribution is used when there are a large number of trials in each of which an event is very unlikely but overall one expects a finite number of events.
Gaussian Distribution
Here we consider the limit of the binomial distribution where N is large and so is Np. Therefore we expect that for P_n to be non-vanishingly small, we need n of the same order as Np
Thus we use Stirling's approximation on binomial distribution to show that

$$s(n) \equiv \ln P_n \simeq n \ln p + (N-n)\ln(1-p) + N \ln N - (N-n)\ln(N-n) - n \ln n$$

for large N, n. It is straightforward to show that

$$\begin{aligned} s'(n) &= \ln p - \ln(1-p) + \ln(N-n) - \ln n \\ s''(n) &= -\frac{1}{N-n} - \frac{1}{n} \end{aligned}$$

You should notice that $s(n)$ is a maximum at $n = Np$ i.e. at the mean value.
Now expand in powers of $x = n - Np$, i.e. make a Taylor expansion around the mean value:

$$s(x) = s(Np) + xs'(Np) + \frac{x^2}{2}s'(Np)\,...$$

Note that since s is maximised at the mean value the first non-zero term in the expansion is in x^2 and

$$s(x) \simeq s(NP) = \frac{x^2}{2N(1-p)p}$$

When we return to $P_n = \exp s(n)$ we find

$$P_n \simeq P_{Np} \exp\left(\frac{-(n-Np)^2}{2N(1-p)p}\right)$$

In order to determine the constant P_{N_p} which serves to normalise the distribution we can make the approximation

$$1 = \sum_{n-0}^{N} P_m \simeq \int_{-\infty}^{\infty} P_{Np} \exp\left(\frac{-(x^2}{2N(1-p)p}\right) dx$$

where extending the limits in this way produces no error since the additional contributions are vanishingly small.

We then invoke a standard integral you should be familiar with

$$\int_{-\infty}^{\infty} \exp\left(\frac{-(x^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}.$$

From this we deduce the correct normalisation for the Gaussian approximation to the binomial distribution as

$$\boxed{P_n \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(n-Np)^2}{2\sigma^2}\right)}$$

where $\sigma^2 = Np(1-p)$.
Note that if $|n - Np| \gg N^{1/2}$ the probability becomes vanishingly small due to the argument of the exponential becoming very large.
The Gaussian distribution is often referred to as a 'Normal distribution'.

1.4 Simplicity at large N: a toy example

Suppose that N coins are tossed; we denote by n the number of heads obtained. We will determine the probability distribution p_n (ie the set of probabilities $p_0, p_1 \ldots, p_N$) characterising the possible outcomes.

Denote by p the probability that a head results from a single toss; then $q \equiv 1 - p$ gives the corresponding probability for a tail.

There are *many distinct ways* of getting n heads from N tosses, differing according to which tosses give heads. It should be clear that each one of these distinct ways has the same probability, namely $p^n q^{N-n}$, and recalling your probability and statistics lectures, the number of distinct ways in which we can get n heads from N tosses is given by the binomial coefficient:

$$\binom{N}{n} \equiv \frac{N!}{(N-n)!n!}$$

It follows that

$$\begin{aligned} p_n &\equiv \text{number of distinct ways of obtaining } n \text{ heads} \\ &\quad \times \text{probability of any specific way of getting } n \text{ heads} \\ &= \binom{N}{n} p^n q^{N-m} \end{aligned}$$

This is the binomial distribution of probabilities. The distribution is characterised by two key parameters, the *mean* $\overline{n}$ and the *variance* $\overline{\Delta}n^2$. Remind yourself from your probability and statistics lectures that $\overline{n} \equiv \sum_{n=0}^{N} np_n = Np$, and $\overline{\Delta n^2} \equiv \sum_{n=0}^{N} (n - \overline{n})^2 p_n = Npq$.

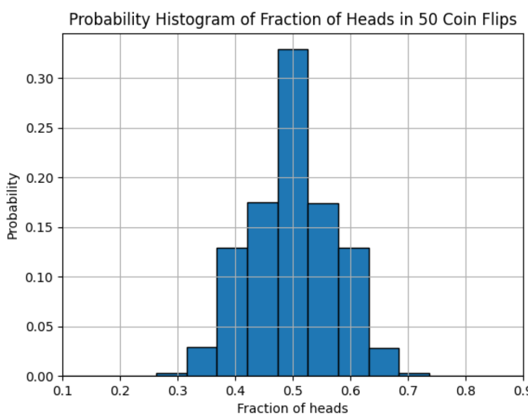
In the present context, setting $p = q = \frac{1}{2}$ (for an unbiased coin) and defining $f \equiv n/N$, the fraction of tosses giving heads, we have

$$\overline{f} = \frac{\overline{n}}{N} = p = \frac{1}{2}$$

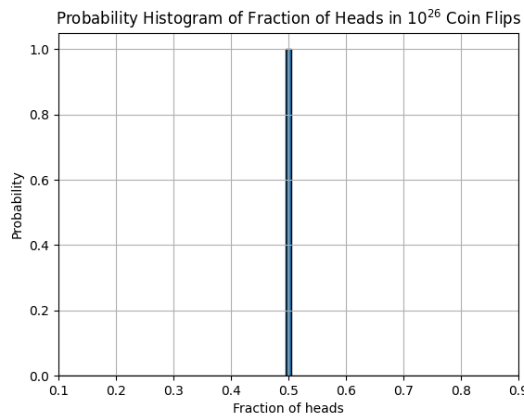
and

$$(\overline{\Delta f^2})^{1/2} \equiv \frac{(\overline{\Delta n^2})^{1/2}}{N} = \left(\frac{pq}{N}\right)^{1/2} = \frac{1}{2N^{1/2}} \tag{1}$$

Equation 1 shows that the typical deviation of f from its mean value is vanishingly small (it is $O(N^{-1/2})$ for N large). Thus for large N we can be very sure that f will always effectively coincide with its mean (cf Figure 1b). The virtual certainty that comes from dealing with large numbers is one of the distinctive features of statistical physics.



(a) For $N = 50$ tosses, we can be *reasonably* sure that f will be close to 0.5



(b) For $N = 10^{26}$ tosses We can be *absolutely* sure that f will be indistinguishable from 0.5

Figure 1: Left: a probability histogram of the fraction of heads f obtained from repeated trials of $N = 50$ coin tosses. Right: for a vary large number $N = 10^{26}$ tosses the histogram narrows essentially to a δ function.

You can investigate for yourself the effect of changing the number of coin tosses N in this python simulation (copy it into your favourite Python runtime environment):

```python
import random
import matplotlib.pyplot as plt
import numpy as np

def flip_coin(n):
    """Simulate flipping a fair coin n times."""
    outcomes = [random.choice(['H', 'T']) for _ in range(n)]
    heads_count = outcomes.count('H')
    return heads_count / n

def simulate_coin_flips(N, M):
    """Simulate flipping a fair coin N times and repeat M times."""
    fractions = [flip_coin(N) for _ in range(M)]
    return fractions

def plot_probability_histogram(fractions):
    """Plot a histogram of the probabilities of the fractions of heads."""
    weights = np.ones_like(fractions) / len(fractions)
    plt.hist(fractions, bins=np.linspace(0, 1.0, 50), weights=weights, edgecolor='black')
    plt.xlabel('Fraction of heads')
    plt.ylabel('Probability')
    plt.xticks(np.arange(0, 1, 0.1))
    plt.xlim(0.1,0.9)
    plt.title('Probability Histogram of Fraction of Heads in {} Coin Flips'.format(N))
    plt.grid(True)
    plt.show()

#if __name__ == "__main__":
    N = int(input("Enter the number of times to flip the coin each time: "))
    M = int(input("Enter the number of times to repeat the simulation: "))
    fractions = simulate_coin_flips(N, M)
    plot_probability_histogram(fractions)
```