PH3012 Thermal and Statistical Physics
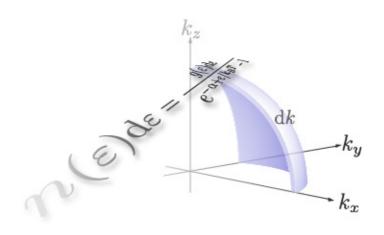
Statistical Physics

Dr Irina Leonhardt

(room 210, il4@st-andrews.ac.uk)

Contents

5 An application of Boltzmann statistics: Paramagnets 31

6 Indistinguishable particles: quantum gases and fluids 37

7 The Fermi-Dirac distribution in more detail 45

8 The behaviour of bosons: photons and black-body radiation 49

9 Conserved bosons and Bose-Einstein condensation 55

10 Gases in the classical limit

Recommended texts

For the statistical physics part of PH3012, the recommended texts are:

Statistical Mechanics: A Survival Guide, by Mike Glazer and Justin Wark, published by Oxford University Press, and

Statistical Physics, 2nd ed., by Tony Guénault, published by Springer.

The former costs about 25 GBP to buy new and the latter about 40 GBP (or you could look for second-hand copies, that are a bit cheaper). There are multiple copies of both in the J. F. Allen Library (Glazer and Wark: QC174.8G6, also as an e-book; Guénault: QC174.8G84).

While I would recommend obtaining your own copy of one of the above books, the course notes and lecture material aim to to provide a good coverage. If you do decide to get your own copy, just pick one! There is a significant overlap in content and they are pitched at pretty much the same level. I think that Guénault (which I would pick) has the edge in the number of examples and covers a wider range, but has some (minor) oddities in notation. The two books tackle the same ideas in a slightly different order, but this doesn't really matter. There are also a few differences in notation, but I'll highlight these as we go to avoid confusion. Glazer & Wark was the main recommended text for this course over the last few years, so there may be more second-hand copies drifting around locally.

For the introductory section on probability, you may wish to consult chapter 1 of Glazer & Wark, or the relevant portions of *Advanced Engineering Mathematics* by Erwin Kreyszig (various editions available in the library, classmark QA401.K8), or indeed many books on probability (classmarks in the region of QA273).

Recommended for the thermal physics part of the course:

Thermal Physics, 2nd Edition, C. B. P. Finn, CRC Press, 1993 (Library classmark QC311.F5F93) This is part of the same book series as Guénault, and (rather handily) they refer to each other at useful points to make clearer the links between statistical physics and thermodynamics.

Further reading:

There are a few more books that you can consult for a better understanding of the concepts and calculations involved in Thermal and Statistical Physics.

- Daniel V. Schroeder " An Introduction to Thermal Physics" (Addison Wesley Longman, 2000)

- F. Mandl "Statistical Physics" (Wiley, 1988)

- Charles Kittel "Thermal Physics" (Wiley, 1969)

Chapter 1

Introduction and thermodynamics review

1.1   What is thermal and statistical physics about?

Mechanics and dynamics are vital to modern physics, but they cannot be used to describe the behaviour of large numbers of particles. (Imagine trying to model the behaviour of a mole of gas by accounting for the individual positions and motions of 6×10^{23} molecules!)

Thermal and statistical physics *can* describe this behaviour and so perhaps could be said to begin the journey into "real world" physics. The development of thermodynamics and statistical mechanics through the 19th and 20th centuries was an intellectual revolution as profound as that of quantum physics. It turns out that one can make remarkably few starting assumptions but *still* come up with a theory about the behaviour of a large system of "particles" that has *enormous* predictive power.

Questions are often asked about how thermodynamics relates to statistical mechanics. Properly constructed thermodynamic laws (including the third law of thermodynamics, which required an appreciation of quantum mechanics) are overarching laws of nature which, as far as we know, must be obeyed by any large system. Thermodynamics is more general than statistical mechanics: fewer assumptions are required to construct the laws of thermodynamics than those of statistical mechanics, making thermodynamics a very elegant subject.

Arguably one of the most abstract areas of physics, the elegance of thermodynamics comes perhaps at the price of easy *physical* understanding. Some concepts can hopefully made clearer by starting with Thermal Physics and then followed by Statistical Physics, although we shall see that we do not completely separate the two (and, indeed, why should we, as they are so closely related).

1.2   Thermodynamics review

Let us begin with a quick look at some key thermodynamics ideas that you have already encountered. Along the way on our statistical journey, we'll need to make some links to thermodynamics to really appreciate what the statistical mechanics is telling us, so it's worth a brief summary.

The Zeroth law of thermodynamics:

The zeroth law of thermodynamics is concerned with the idea of *thermal equilibrium*. Consider three thermodynamic systems, which we'll call A, B and C. If A is in thermal equilibrium with B, and B is in thermal equilibrium with C, then A is also in thermal equilibrium with C. This actually forms the whole basis of the idea of *temperature*. (see Finn, chapter 1)

The First law of thermodynamics:

The first law of thermodynamics essentially tells us about the conservation of energy. The change in internal energy of a system is the heat input to the system plus the work done on the system, i.e.

$$\Delta U = Q_{in} + W_{on} \tag{1.1}$$

where U denotes the internal energy of the system, W is the work done *on* it and Q is the heat *input* to the system. (If we were talking about the work done *by* the system, or the heat that is *output* from the system, then the sign(s) of W and Q would need to change, but we usually just stick with the sign convention implied by equation 1.1 and then let the sign that we work out for W or Q tell us whether there is net work done on or by the system, and whether the net heat comes into, or goes out of, the system.)

Remember, if our system changes from some definite state, A, to some other definite state, B, then ΔU is determined (from U_A and U_B), so obviously the quantity $(Q + W)$ is determined, but Q and W are not *separately* determined, as both Q and W depend upon *how* we get from state A to state B. (For the special case of **adiabatic** processes, i.e. where the system we are considering is thermally isolated, then $Q = 0$ by definition, in which case W *is* independent of the path taken from state A to B, but this is not the case in general.) ΔU is an example of a *state function*, the defining property of which is that it doesn't matter how the state (defined, for example, by measurable variables such as P, V and T) was reached. Q and W, on the other hand, do depend on how we got from state A to state B, so are not state functions.

It may come as some comfort to learn that statistical mechanics helps us form an intuitive picture of the difference between *heat* and *work*.

The Second law of thermodynamics:

While the first law is concerned with the conservation (i.e. balance of) energy in physical processes, the *second* law of thermodynamics is concerned with the *direction* of such processes. A simple example is that heat will never flow from a colder body to a hotter body without some kind of intervention. (A fridge is an example where we can make heat from the cold interior to the warmer heat exchanger at the back, but this requires some mechanism and energy input, it would never occur *spontaneously*.) **Entropy** (usually denoted by S) was introduced to allow this notion of the direction of a process to be quantified. In any *isolated* system, the total net entropy can at best remain the same or increase, it can never decrease. One of the good things about statistical physics is that the notion of what entropy means is perhaps less abstract than in thermodynamics.

If an amount of heat, ΔQ is input to a system at temperature T, then the associated change of entropy is given by

$$\Delta S = \frac{\Delta Q}{T} \tag{1.2}$$

and in any isolated system, the total net entropy either stays the same or increases.

The Third law of thermodynamics:

As $T \to 0$, $S \to 0$, a quantum mechanical effect. We'll see some justification of this as we proceed.

The thermodynamic identity

One of the most important equations in thermodynamics:

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V \tag{1.3}$$

This applies to a system in which the amount of substance doesn't change $(\mathrm{d}N = 0)$.

Enthalpy

The **enthalpy** of a system is defined as

$$H = U + PV \tag{1.4}$$

For an infinitesimal processes, we can express the infinitesimal change in the enthalpy as

$$\mathrm{d}H = \mathrm{d}U + P\mathrm{d}V + V\mathrm{d}P \tag{1.5}$$

which, along with the first law and a wee bit of maths tells us that for the *particular* case of a reversible process at constant pressure $(\mathrm{d}P = 0)$,

$$\mathrm{d}H = \mathrm{d}Q \tag{1.6}$$

Gibbs free energy

The **Gibbs free energy** (also known as the *Gibbs function*) is given by

$$G = H - TS = U - TS + PV \tag{1.7}$$

Helmholtz free energy

The **Helmholtz free energy** (also known as the *Helmholtz function*) is defined as

$$F = U - TS \tag{1.8}$$

Sometimes, this is denoted by A, rather than F, but we'll stick with F in this course. For an infinitesimal processes, we can express the infinitesimal change in the Helmholtz free energy as

$$\mathrm{d}F = \mathrm{d}U - T\mathrm{d}S - S\mathrm{d}T = -S\mathrm{d}T - P\mathrm{d}V \tag{1.9}$$

From which in turn we can derive a variety of formulas for partial derivatives. For instance, holding V and N fixed yields the identity

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,N} \tag{1.10}$$

The Helmholtz function in particular allows an important and straightforward link to be made between thermodynamics and statistical mechanics, which is one of the key motivations for reviewing your existing thermodynamics knowledge at this point.

Chapter 2

Introduction to probability

We'll now have an introduction to some of the key ideas on probability. This may initially seem like a bit of a non-sequitur, but its utility will become apparent almost immediately. Some of the ideas here also relate to (and indeed, may be familiar from) quantum mechanics, and others are also important in the consideration of the uncertainty in experiment results. The more that you can link the use of a particular mathematical tool across different subjects in physics, the more intuitive their application to 'new' problems will become.

2.1 The probability of an event and the probability distribution

The *probability* of an event is just the chance that it will happen. We can think of a procedure or experiment which has r mutually exclusive outcomes (i.e., **only one** of them can happen) and the outcomes are r values of some variable x, represented by x_i $(i = 1 \dots r)$. (For example, if we throw a normal die, then there are $r = 6$ outcomes, and x can take the values 1 to 6 inclusive.)

We can define p_i as the probability of outcome i, so p_i is the probability that the outcome of our procedure/measurement is x_i. If we know the complete set p_i, $(i = 1 \dots r)$, then we have a *complete* statistical description of the problem. p_i defines the *probability distribution* of the *discrete* variable x, and is illustrated in figure 2.1.

A key property of the probability distribution is that

$$\sum_i p_i = 1 \tag{2.1}$$

This makes perfect sense as it is obvious that if we make a measurement, we *must* measure one of the possible outcomes x_i. It also means that the probability of the outcome *not* being x_i is simply $1 - p_i$.

We denote the *mean* value of x by $\langle x \rangle$:

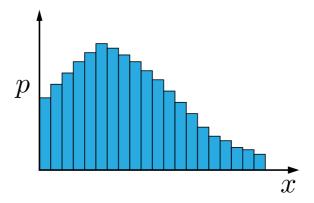$$\langle x \rangle = \sum_i p_i x_i \tag{2.2}$$

Figure 2.1: A completely made-up example of a probability distribution of a discrete variable.

The **variance** is a measure of the spread, or width of the distribution and is given by

$$\left\langle \Delta x^2 \right\rangle = \sum_i p_i \left(x_i - \left\langle x \right\rangle \right)^2 \tag{2.3}$$

A related quantity is the *standard deviation*, σ:

$$\sigma = \sqrt{\left\langle \Delta x^2 \right\rangle} \tag{2.4}$$

(This is a handy quantity, as it obviously has the same dimension as whatever x is measured in.)

If we have a function of x, then the mean value is given by

$$\left\langle f(x) \right\rangle = \sum_i p_i f(x_i) \tag{2.5}$$

 A classic result of elementary statistics is that we can express the variance, $\sigma^2 = \left\langle \Delta x^2 \right\rangle$, as the "mean of the squares minus the square of the means":

$$\begin{aligned}
\sigma^2 &= \sum_i p_i \left(x_i - \left\langle x \right\rangle \right)^2 \\
&= \sum_i p_i \left(x_i^2 - 2 x_i \left\langle x \right\rangle + \left\langle x \right\rangle^2 \right) \\
&= \left\langle x^2 \right\rangle - 2 \left\langle x \right\rangle^2 + \left\langle x \right\rangle^2 \\
&= \left\langle x^2 \right\rangle - \left\langle x \right\rangle^2
\end{aligned} \tag{2.6}$$

What about *continuous* variables? Consider a procedure whose outcome is represented by a variable x which can take *any* value in a given range. Imagine chopping up the range into a series of intervals $x_1 \rightarrow x_2$, $x_2 \rightarrow x_3$ and so on, with each small interval having the *same* width δx. We now define $p_{x_i \rightarrow x_i + \delta x}$ as the probability of getting an outcome in the interval $x_i \rightarrow x_i + \delta x$. This probability must be proportional to δx, with proportionality constant p which depends on the particular interval, i.e. $p = p(x_i)$, so $p_{x_i \rightarrow x_i + \delta x} = p(x_i) \delta x$ and

$$p(x_i) = \lim_{\delta x \rightarrow 0} \frac{p_{x_i \rightarrow x_i + \delta x}}{\delta x} \tag{2.7}$$

$p(x)$ is defined as *the probability density* of the continuous variable x and is best envisaged just as a continuous probability distribution, in which we need to be careful about normalisation and such.

2.2   Defining characteristics of the probability density function

$p(x)\mathrm{d}x$ is the probability of a result in the range $x \to x + \mathrm{d}x$. The probability of a result being in the "macroscopic' range $a \to b$ is given by:

$$p_{a \to b} = \int_a^b p(x)\mathrm{d}x \tag{2.8}$$

and the condition of *normalisation* gives

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1 \tag{2.9}$$

Because of the normalisation condition, then if x has units, the units of $p(x)$ will be the inverse of the units of x. Similarly to the case for a discrete probability distribution, the average value of a function $f(x)$ will be given by

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x \tag{2.10}$$

The mean and the variance are defined using equation 2.10 and the definitions given for discrete probability distributions.

2.3   Independent events

If we have two events that are *uncorrelated*, or completely independent, then the probability that they both occur is simply the product of the individual probabilities, so if we have two events A and B, with respective probabilities p_A and p_B, then the probability that they both occur is

$$p_{A,B} = p_A \, p_B \tag{2.11}$$

The above can be extended to more than two events very simply. If we have three independent events A, B and C, for example, then the probability of all three occurring is $p_A \, p_B \, p_C$.

If a set of experiments shows that the above is not the case, then the events are not independent, i.e. they are *correlated*.

2.4   The 'normal' or 'Gaussian' distribution

This is one of the most widely studied distributions. Imagine that we are measuring a quantity whose *true* value is $\langle x \rangle$. Our measurement process is not perfect (due to, say, random errors in an imperfect setup, or perhaps due to intrinsic uncertainties like those in quantum theory). In this case,

the probability density function associated with our process will be a very good approximation to a *Gaussian*:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \langle x \rangle)^2}{2\,\sigma^2}\right) \tag{2.12}$$

The factor in front of the exponential ensures that the distribution is normalised. The full width of the peak at half the maximum value is $2.355\,\sigma$.

2.5   The theory of sampling*

*This material is not examinable.

If we knew the probability distribution (known as the *parent distribution*), then life would be easy. However, we almost always do not. What we will have instead is a set of N measurements or observations that we can plot to form a *sample* distribution, which will be a pretty poor approximation to the parent unless N is very large. However, it gives a much better approximation to the true mean (i.e. the mean of the undetermined parent distribution) and a rational quantitative estimate of how much better an approximation it really is.

We denote our N independent measurements of x as $x_1 \ldots x_N$. We denote the mean of this sample as $\overline{x}$ and the variance of the sample as s^2, and define them as

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.13}$$

and

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 = \overline{x^2} - \overline{x}^2 \tag{2.14}$$

The notation is important: a bar ($^-$) above a symbol denotes an average over a sample, and $\langle\ \rangle$ denotes the theoretical average over the notional infinite parent population that defines the parent distribution, i.e.

$$\langle x \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.15}$$

Given our sample, $\bar{x}$ is our most rational estimate of $\langle x \rangle$.

Now let's look at the standard deviation:

$$\langle s^2 \rangle = \langle \overline{x^2} \rangle - \langle \bar{x}^2 \rangle \tag{2.16}$$

and we examine the two terms on the right hand side in turn.

$$\langle \overline{x^2} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right\rangle$$

$$= \frac{1}{N} \sum_{i=1}^{N} \langle x_i^2 \rangle$$

$$= \langle x^2 \rangle \tag{2.17}$$

which is straightforward. The 2nd term, however, is not so simple:

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

$$= \frac{1}{N^2} \sum_{i,j=1}^{N} \langle x_i x_j \rangle$$

$$= \frac{1}{N^2} \sum_{(i=j)=1}^{N} \langle x_i^2 \rangle + \frac{1}{N^2} \sum_{i,j=1,i\neq j}^{N} \langle x_i x_j \rangle$$

The significance of the second term in the last step above is that if x_i and x_j are uncorrelated (which they are if $i \neq j$, then $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle = \langle x \rangle^2$, so

$$\langle \overline{x}^2 \rangle = \frac{1}{N} \langle x^2 \rangle + \frac{N(N-1)}{N^2} \langle x \rangle^2 \tag{2.18}$$

If we substitute the results of equations 2.18 and 2.17 into equation 2.16 then we end up with

$$\langle s^2 \rangle = \left(1 - \frac{1}{N} \right) \left(\langle x^2 \rangle - \langle x \rangle^2 \right) = \frac{N-1}{N} \sigma^2 \tag{2.19}$$

This means that based on our knowledge of our single sample, our most rational estimate of σ^2 is not s^2 but

$$\sigma^2 \approx \frac{N}{N-1} s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{2.20}$$

If N is more than a few, then our most rational estimate of σ^2 becomes practically equivalent to just using s^2.

2.6   The error in the mean*

*This material is not examinable.

In many cases of practical importance such as an experimental measurement subject to random error, the aim of the exercise is to make as good an estimate as we can of the value that we are trying to measure. The value we are trying to measure will be the mean of the *parent* distribution. As previously stated, our most rational estimate of the "true" or "parent" mean, $\langle x \rangle$, is our sample mean, $\overline{x}$. Arguably the most powerful aspect of sampling theory is the ability it gives us to *estimate the error, or uncertainty, in our estimate of* $\overline{x}$.

We want to make the most rational estimate that we can of the deviation of $\overline{x}$ from $\langle x \rangle$. By analogy to what's gone before, we define the *standard error in the mean*, σ_m:

$$\sigma_m^2 = \left\langle \left(\overline{x} - \langle x \rangle \right)^2 \right\rangle \tag{2.21}$$

It is important to clearly understand the meaning of equation 2.21: it defines σ_m as the standard deviation that would result if we carried out a very large number of samples, and plotted the distributions

of their means. This is an extremely useful idea.

$$
\begin{aligned}
\sigma_m^2 &= \left\langle \left(\overline{x} - \langle x \rangle \right)^2 \right\rangle \\
&= \left\langle \overline{x}^2 - 2\overline{x}\langle x \rangle + \langle x \rangle^2 \right\rangle \\
&= \frac{1}{N}\langle x^2 \rangle + \left(1 - \frac{1}{N} \right)\langle x \rangle^2 + \langle x \rangle^2 - 2\langle x \rangle^2 \quad \text{using equation 2.18} \\
&= \frac{1}{N}\left(\langle x^2 \rangle - \langle x \rangle^2 \right) \\
&= \frac{\sigma^2}{N}
\end{aligned}
\tag{2.22}
$$

This leads to the *extremely* important result

$$
\sigma_m = \frac{\sigma}{\sqrt{N}}
\tag{2.23}
$$

This result may look simple but is absolutely core to experimental and statistical physics. It says that if we take a sample of measurements (or in statistical physics, a sample assembly of particles) the statistical precision with which we know the mean (or other average properties) of that sample improves as the square root of the number of measurements or particles over which we take that average.

Please note that this powerful result does rely on some assumptions: in particular, it assumes that the measurements or particles in the sample are *random* and *independent*: before using this result, you will have to assess the extent to which this assumption is true.

Chapter 3

From probability to statistical physics

Having described where we are going, we'll have another look at probability theory, but this time we have the motivation of introducing and describing some of the language of statistical mechanics.

Let's consider a game in which we toss four coins. Each coin toss has two possible outcomes (heads or tails, denoted H or T), so the game has a total of $2^4 = 16$ outcomes (e.g. HTHH, TTTH, etc). In statistical physics, each single possible outcome for the 4 toss game is referred to as a **microstate**. A microstate is a *complete* description of the results of the 4-coin system after they have all been tossed, in that we know *for each coin* whether it's come up heads or tails.[1] In our game, if we assume fair coins, then each 4-coin outcome, i.e. *each microstate*, is equally likely.

In our game, it is simple to list each of the individual microstates, as there are only four coins. There are two ways we could describe the outcome of our game: we could, for example say that the state of the system was (say) THHT: this would be a complete description of the state of the system, so we are identifying a particular microstate. Alternatively, could just say that the outcome was two heads and two tails. While clearly a less complete description (as it tells us nothing about *which* of the coins are the two heads and which are the two tails), this classification can still be useful.

---

[1]The description is complete as we are assuming that the side that comes up for each coin is the only thing we care about, i.e. we are not interested in its motion or position, or anything like that....
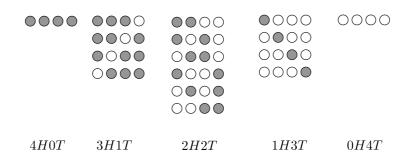

Figure 3.1: The sixteen possible outcomes (microstates) from our four coin-toss game, grouped into macrostates defined by the number of heads or tails.

In statistical physics, such a less complete description of a system is referred to as a **macrostate**. When considering large systems, macrostates can be more useful: imagine trying to write out all the microstates possible for a game or system in which instead of four coins, we had 10^{23} coins. (In fact, even if you imagine only ten coins, you're going to get bored pretty quickly!) Figure 3.1 shows all the possible microstates for the 4 coin toss game, grouped into macrostates, where the macrostate is determined by the total number of heads (or tails).[2]

We will have macrostates with four heads and no tails (which we'll call 4H0T for short), 3H1T, 2H2T, 1H3T, 1H4T, so 5 macrostates in total. (We can simplify our description a bit further: as each coin can only come up heads or tails, then for a given number of heads, the number of tails is fixed, so we only need to specify the number of heads, or specify the number of tails. If we choose the former, then we could label the the macrostates as 4H, 3H, 2H, 1H and 0H.)

If we think about a "thermodynamically large" system with a massive number of particles, then there is going to still be a huge number of macrostates, so things may still look pretty complicated. However, note that the most probable macrostate is also the one containing the mean number of heads and tails. (We know that chances of arriving at each microstate are equal, i.e. that the probability of each 4-coin combination (microstate) arising is the same. Since there are more ways to arrive at 2H than at any other particular macrostate, and each way (microstate) has the same probability, then the probability of ending up in macrostate 2H must be larger than that of ending up in any other particular macrostate.)

3.1 Factorial notation and combinatorics

Imagine that we have three different objects, called A, B and C. We can arrange these objects in different ways: ABC, ACB, BAC, BCA, CAB, CBA. There is a total of 6 arrangements, and $6 = 3 \times 2 \times 1 = 3!$ The exclamation marks denotes "factorial", which is defined so that for any *positive integer n*,

$$n! = n \times (n-1) \times (n-2) \times (n-3) \times \ldots \times 2 \times 1 \tag{3.1}$$

You can work out for yourself that this function gets large pretty quickly! Let's consider our three objects again, and think about how we could arrange any two out of the three: AB, BA, AC, CA, BC, CB. Again, we get six ways. However, *if we are not concerned about the order* then we only have 3 possibilities, as AC is equivalent to CA, etc. In the general case, the number, C, of **combinations** of n objects taken from a set of $N > n$ is given by

$$C = \frac{N!}{n!\,(N-n)!} \tag{3.2}$$

(Note that the term "combination" applies to the case where the order is not important. If the order were important, we would be talking about a *permutation*: in general, combinations are concerned with the choice of objects, and permutations with the arrangement of objects.)

Example 1: The number of combinations of 3 objects chosen from 4 is

$$\frac{4!}{3!1!} = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} = 4$$

---

[2]We are assuming that the chances of a coin balancing on its edge is zero...

Example 2:   The number of combinations of 2 objects chosen from 4 is

$$\frac{4!}{2!2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = 6 \tag{3.3}$$

If we look back at the discussion of microstates and macrostates we had based around our 4 coin-toss game, note that for that special example our combinations formula is calculating the number of microstates in a given macrostate. We shall see in time that this is not a coincidence. (The general statistical physics formula for calculating the number of microstates in a macrostate reduces to (3.2) in our particular example. If you work out the number of combinations of 3 objects chosen from 6, you'll realise why using six coins for the example would make a figure showing all the possible 'microstates' too big for the page!

3.2   Realistically large numbers

Let's extend our consideration to large numbers of coins. "All" we need to do is calculate the number of combinations of, for example, things like n heads and n tails for $2n \sim 10^{23}$. (Off you go then...) In fact, to see the difficulty involved, imagine we had only ten coins: it would still be a really tedious exercise to figure out all the microstates.

However, there is a way out of this: for large n, we can use *Stirling's approximation*:

$$\ln n! \approx n \ln n - n \tag{3.4}$$

A graphical proof of this is provided in Guénault (appendix B), and Glazer & Wark (pages 5–6).

Now let's use Stirling's approximation to calculate the occupation of the most probable macrostate. We denote the occupation number of a macrostate by Ω. *(Note: Guénault uses Ω to denote the* **total** *number of microstates (i.e. over more than one macrostate), and t to denote the number of microstates in a given macrostate. When we get a bit further on, we'll end up using an assumption that lessens the importance of this distinction, but you should be aware of it if you consult the textbook.)* For a large ensemble of two-state systems the occupation of a macrostate containing n "heads" and $N - n$ "tails" is

$$\Omega = \frac{N!}{n! \, (N - n)!} \tag{3.5}$$

We handle this using Stirling's approximation:

$$\ln \Omega = \ln N! - \ln n! - \ln(N - n)!$$
$$\approx N \ln N - N - n \ln n + n - (N - n) \ln(N - n) + N - n$$

Stirling's approximation for $\ln n!$ is only valid for large n, but given that a mole of a substance contains the order of 10^{23} particles, we have no trouble justifying its use!

The reason that Stirling's approximation is relevant here is that we are looking for the maximum in Ω. As the logarithm function $\ln x$ is a monotonically increasing function of x, then the maximum in $\ln x$ will occur when x is a maximum. Thus, we look for a maximum in $\ln \Omega$, and to do so begin by
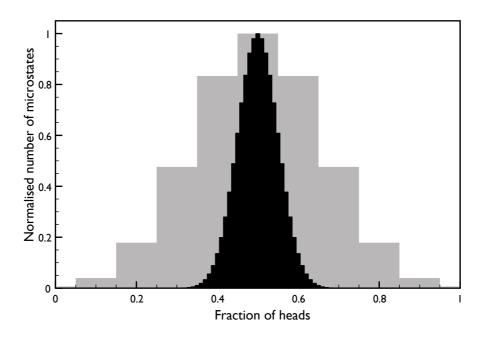
Figure 3.2: The normalised number of microstates for a multiple coin system as a function of the fraction of coins that come up as heads. The black histogram is for 100 coins and the grey histogram is for 10 coins.

attacking it with Stirling's approximation:

$$\frac{d \ln \Omega}{dn} = -\ln n + \ln(N - n)$$

$$= \ln\left(\frac{(N - n)}{n}\right)$$

For this to be zero, $(N - n)/n$ must equal one, so $N - n = n$ and $n = N/2$.

Figure 3.2 shows normalised histograms of the number of microstates as a function of the fraction of coins that come up as heads, for two systems of 10 coins and 100 coins. As N gets larger and larger, the full width at half maximum of the peak for a system of N coins will approach $\sqrt{N}$. However, the position of the maximum of Ω will always be at $N/2$, so the fractional width (i.e. the ratio of the width to the position of the maximum) goes as $1/\sqrt{N}$. (For 10^{23} coins, the fractional width would be about $3 \times 10^{-10}\%$!)

3.3   Macrostates and microstates

In the coins example above, the definitions of macrostate and microstate are pretty straightforward: the former is the total numbers of heads and tails (which can be derived from the total number of heads or the total number of tails) and the latter is a coin-by coin definition of whether it has come up heads or tails.

Real physical systems of the kind we are interested in aren't made up of large numbers of coins, so we need to think about what corresponds to a macrostate of a real system. We are going to

almost entirely restrict ourselves to isolated systems of a single type of particle in which the number of particles in the system (N), the volume of the system (V) and the internal energy of the system (U) are all fixed. Some texts (Guénault included) use this description for the term 'macrostate', but here we'll use the term to mean something a bit more specific.

Imagine that we have a (very small!) crystal of 8 atoms, each of which can have the same set of energies, 0, $\mathcal{E}$, $2\mathcal{E}$, etc., and the total energy of our crystal (which must be the sum of the energies of each atom) is $6\mathcal{E}$. As we are talking about a crystal, the positions are fixed, so we can tell the atoms apart by their position.

In this case, a *macrostate* is a description of how many atoms have a particular energy, and a *microstate* is the complete description of the energy that each atom has. We are choosing to define 'macrostate' as meaning a particular distribution of particles in energy levels.

One possible macrostate is where one atom has energy $6\mathcal{E}$ and the rest have energy 0. As we have 8 atoms, and we can tell them apart, then there must be 8 ways of doing this, so the number of microstates in this macrostate is simply 8.

Another possible macrostate is to have two atoms with energy $3\mathcal{E}$, and the rest have energy 0. Obviously (hopefully!), there are more ways of doing this (i.e. microstates) than for the above case. In this case, we are selecting combinations of 2 particles from 8, so the number of ways of doing this will be $8!/(6! \times 2!) = 28$. Thus, there are more microstates in this macrostate, so this macrostate is more likely than the first case. (The most likely macrostate has one atom each with energy $\mathcal{E}$, $2\mathcal{E}$ and $3\mathcal{E}$.)

We are dealing with a tiny number of atoms here. However, as the number of atoms is increased, the relative probability of the most probable macrostate increases. For very large particle numbers, the system is overwhelmingly likely to be in a state close to the most probable macrostate, as in the coins example.

3.4   Basic postulates of statistical mechanics

One assumption, on which statistical physics ultimately rests, is that for a system in thermodynamic equilibrium all the microstates that are consistent with the constraints on the system are *equally probable*.
This is quite a remarkable statement, and this assumption is central to the operation of statistical physics. We'll take it on trust, but Guénault chapter one offers some intuitive justification (although remember that it does remain an assumption!) One justification mentioned by Mandl is, that if all we know about a system is the specification of its macrostate, then we need to make some assumption about the probabilities of each of the microstates that would be consistent with the macrostate. We have no reason to regard one microstate as being preferred to any other, so we must assign equal probabilities to them.

Another key postulate is, that the macrostate that is observed for a system is the one that contains the most microstates.
Can we justify this in light of the 4-coin or 8-atom examples above?

Luckily, the answer is yes: 4 or 8 is a tiny number compared to the number of particles one would encounter in a real system. When the particle number is very large, then the probability distribution is much more tightly peaked (as in the coins example).

3.5   Statistical definition of entropy

We will also make considerable use of the statistical definition of entropy:

$$S = k_B \ln \Omega \tag{3.6}$$

S is the entropy as derived by mechanical engineers of the 19th century, and Ω is the number of accessible microstates to the system. We are not going to prove this, but we can at least show that it makes sense. If we have two systems, in which the numbers of microstates are Ω_1 and Ω_2 respectively, then the total number of microstates for the *combined* system must be the product of the two.

However, thermodynamics tells us that the total entropy must be the *sum* of the entropies of the two systems. The logarithm is the only function which makes $S = S_1 + S_2$ consistent with $\Omega = \Omega_1 \Omega_2$.

Equation 3.6 provides a key link and is often called the **'bridge equation'** between statistical mechanics and thermodynamics. This equation will help us considerably when working out the most probable distribution for a system.