PC2135: Thermodynamics and Statistical Physics

Lecturer: Angeline Shu

Semester 1, AY2526

Contents

How to Utilize these Lecture Notes

Our aim in this course is to develop a description of the macroscopic properties of a system via a fundamental understanding of the microscopic behaviour of particles in a *systematic manner.*

> As such, at the start of every chapter, the following points will be encapsulated in a little box as such:
>
> - Overview of the chapter
> - Intended learning outcome
> - Textbook reference
> - Flow of chapter

To ensure that you have fully understood the concepts in the lectures, do ensure you are able to achieve the intended learning outcome of the chapter, and can see the link between the contents in the chapter to the overview of the chapter.

These lecture notes were developed to closely follow the reference I utilized, and each chapter will state the particular sections of the book it was based off. I strongly encourage you to read the Reif's book, as the explanations therein will give you a deeper and more nuanced understanding that the lectures can ever give. Learning concepts take time – struggling with words in the textbook is one of the best ways to learn. I will also introduce references of an alternate book (where possible). Different references have different approaches of explaining certain topics. Whilst reading a different references might cause more confusion at the start, for a more advanced student, these different approaches can be highly enriching. Where possible, I encourage students to look up the supplementary text if they have the capacity.

The main reference this lecture notes are based off will be:
- Reif, Frederick. *Statistical Physics.* Vol. 5 of Berkeley Physics Course, McGraw-Hill, 1967.

The following supplementary references may be useful:
- Schroeder, Daniel V. *An Introduction to Thermal Physics.* Addison-Wesley (Pearson), 1999

- Stowe, Keith. *An Introduction to Thermodynamics and Statistical Mechanics.* 2nd ed., Cambridge University Press, 2007.

Developing a description of the macroscopic properties by drawling links to the microscopic behaviour is no easy feat. One can easily get lost without concrete and explicit examples to anchor the discussion.

> **Illustration I**
>
> As such, it is imperative for students to be aware of the system we are using to illustrate our explanations. The system in question will often be encapsulated in a box like this.

I will draw attention to important definitions and concepts via boxes as below.

> Important definitions and concepts will be in these coloured boxes!

Summaries of our long discussions/derivations or worked examples will be encapsulated in a box too!

> This box will highlight the summary of discussion and derivations as well as worked examples.

And finally, comments will be in these boxes.

> **Box 1: Comments**
>
> Comments that are potentially tangentially related to the course, or that are related but will be properly introduced and discussed later, will be separated from the main text as such.

I believe strongly that no matter how clear a lecturer is, a student really only learns when they connect the materials on their own and form their own framework of understanding the material. My job is to structure the presentation of the material in a way that the it flows smoothly and naturally, and to present this structure in a lucid and comprehensible manner. Your job is to struggle with the concepts and abstract out the main points and connect the ideas in a way that is *your own*. Therefore, part and parcel of using this lecture notes include making your own summary after every chapter/part.

Finally, I do recognize it's no longer "fashionable" nor "trendy" to have content delivered in this form. Shorts, Reels, Snippets and Bite-sized content are all the rage now. Unfortunately, I believe that many nuances get lost in these form of microcontent. Furthermore, students exercise certain skills when processing longer-form content, which are essential competencies tertiary education is meant to cultivate. Therefore, my lectures will be based off these long notes.

Part I

Introduction & Basics

Chapter 1

Macroscopic Systems

Overview:
In attempting to describe a macroscopic system via our understanding of the microscopic characteristics of the system, we build a basic intuition for the concepts of fluctuations, equilibrium and irreversibility. The discussion in this chapter are qualitative in nature. We will be revisiting the concepts explored in this chapter several times throughout this course, sharpening our understanding of said concepts.

Intended Learning Outcomes:
At the end of this chapter, students should be able to:
1. *explain* the fundamental characteristic approach in dealing with macroscopic system.
2. briefly *describe* the concept of equilibrium by considering how a system evolves in time.
3. briefly *describe* the concept of irreversibility.
4. *explain* the meaning of the terms macrostate, microstate and macroscopic parameter with the aid of an example.

Textbook Reference:
Reif: Section 1.1, 1.2, 1.4 and 1.7

Flow of chapter:
1. Fundamental Characteristic Approach
2. Description of Small and Large Systems
   - Illustration I
   - Small systems: Probabilities and Fluctuations
   - Large systems: Probabilities and Fluctuations
3. Approach to Equilibriums and Irreversibility
4. The Equilibrium Scenario

1.1 Fundamental Characteristic Approach

Scientific progress has led us to understand that:

- All matter consists of molecules, built up of atoms. There is overwhelming experimental evidence in support of atomic theory.
- The dynamical behaviour of these atoms and molecules are governed by laws of quantum mechanics, which are well established.
- The electromagnetic forces responsible for the interactions between these atomic particles are very well understood[a].

---
[a]These are generally the only forces relevant. Gravitational forces on atomic particles are generally negligibly small compared to electromagnetic forces. Nuclear forces are not necessary since we do not concern ourselves with the constituents within the atom.

Hence, in principle, we ought to be able to deduce the properties of any macroscopic system from a knowledge of its microscopic constituents.

The problem: A typical macroscopic system contains about 10^{24} interacting atoms! Hence our knowledge is useless in prediction of any sort unless we develop methods to deal with this complexity. These method should allow us to:

- identify the macroscopic parameters most useful in describing the macroscopic system
- quantitatively predict properties of the system

The solution: We find that the absurdly large number of particles results in statistical methods being remarkably effective to assist us in prediction!

1.2 Description of Small and Large Systems

In this section, we will utilize an example to qualitatively demonstrate how absurdly large numbers mess with our intuition, and how the statistics of large numbers could simplify the macroscopic description of a system.

1.2.1 Set-up & Small Systems

> **Illustration I**
>
> Consider N identical (assumed distinguishable[a]) molecules in a container that can be partitioned into two equal parts, with n is the number of molecules in the left half, n' the number of molecules in the right half, such that $n + n' = N$.
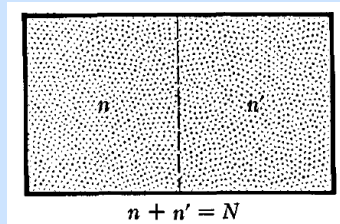>
>
>
> Figure 1.1: A box containing N molecules of ideal gas, subdivided into two equal parts by an imaginary partition, with n molecules on the left and n' molecules on the right
>
> If the gas is sufficiently *dilute*[b], to the point that they are considered *ideal*, they behave like free particles (i.e. no external force exerted on them), only rarely coming into contact with other particles or the wall to interact/collide with them. We will further assume this system is *isolated*, and has been left undisturbed for a very long time.
>
> > **Ideal**: A system that is sufficiently dilute that the interaction between its molecules is almost negligible.
> >
> > **Isolated**: A system that does not interact *at all* with any other system.
>
> For the purposes of this illustration, imagine that we can observe the gas molecules without affecting their motion (utilizing a suitable camera, for instance), and we could capture subsequent frames that show the positions of these molecules at regular time intervals τ_0 that are short. These frames can be thought of as a movie strip.
>
> ---
>
> [a]In Physics, the term *identical particles* implies that the particles are *indistinguishable*, a concept that is central to quantum theory. In the classical treatment, particles are never quite indistinguishable as in the quantum sense. We will discuss this concept of identical particles towards the end of this course. This footnote here is just to clarify (especially for those who know more) what I mean by identical when used here – I mean it in the classical sense as different particles that are in principle distinguishable, but are identical – like identical twins!
>
> [b]Dilute is taken to mean number of molecules per unit volume is small, causing average separation between molecules to be large, and their mutual interactions to be correspondingly small.

Consider the positions of these molecules. A microscopic description or *microstate* of this system would detail the position of every single N molecule within that box. With our knowledge of mechanics, we can indeed predict the location of every single molecule at an arbitrary time t given the initial condition. However, one is rarely interested in this complicated picture.

We would prefer a sufficiently adequate macroscopic description or *macrostate* of the system, which captures sufficient information and for us to make predictions about the system. This macroscopic description would depend on a particular choice *macroscopic parameter*. A macroscopic parameter is a parameter which can be determined by large-scale measurements. Inherent in this parameter is the sensitivity or precision to which our measurements have. For instance, in this example, a macroscopic parameter would be how many particles are in the left half of the box. Depending on the overall number N however, our precision would differ. For $N = 50$, I might consider a ± 5 in my description and a macrostate can be 20-30 particles on the left half of the box. For $N = 10^{24}$ however, a precision of ± 5 would be ridiculous.

> ***Macroscopic parameter***: A parameter which can be determined by large-scale measurements and which describes the macroscopic state of a system. Inherent in this choice is the precision which our measurement will have.
>
> ***Macrostate***: State of the system described without attention to the microscopic details, by specifying only the quantities which can be determined by macroscopic measurements (i.e. macroscopic parameter)
>
> ***Microstate***: State of a system described in microscopic detail by the most complete specification, according to the laws of mechanics

Probability and Configurations

If $N = 2$, we would have four different *configurations*, where configuration in this context would be each distinct way in which a molecules can be distributed between the two halves.

Your Illustration: Four different configuration of $N = 2$

> ***Configuration***: A particular arrangement that would match a chosen macroscopic constraint (in our illustration, the constraint is the number of molecules N)

With $N = 3$, each molecule will have two possible locations, and the total possible configuration is equal to $2 \times 2 \times 2 = 2^3 = 8$. With $N = 4$, we would have 16 configurations. We list them below. Note that C_n here represents number of configuration with n molecules on left, and P_n represents the probability of having n molecules on the left

1	2	3	4	n	n'	$C(n)$	P_n
L	L	L	L	4	0	1	$1/16$
L	L	L	R				
L	L	R	L				
L	R	L	L	3	1	4	$\dfrac{4}{16} = \dfrac{1}{4}$
R	L	L	L				
L	L	R	R				
L	R	L	R				
R	L	L	R				$\dfrac{6}{16} = \dfrac{3}{8}$
L	R	R	L	2	2	6	
R	L	R	L				
R	R	L	L				
L	R	R	R				
R	L	R	R				
R	R	L	R	1	3	4	$\dfrac{4}{16} = \dfrac{1}{4}$
R	R	R	L				
R	R	R	R	0	4	1	$1/16$

Table 1.1: Possible configurations and probabilities for when $N = 4$

Due to the nature of these molecules that are moving about and colliding with each other, we're never quite sure how many molecules will be in the left half of the box due to *fluctuations*, but we can gauge the chances of expecting n molecules there. The probability of the system being in any one of the configurations about would be 1/16, as there are 16 possible configurations and we assume *each of them being equally likely* (see Box 1.1). The probability of then having n molecules on the left side is given by the last column in the table.

> **Box 1.1: Fundamental Postulate of Statistical Mechanics**
>
> I have found that although people have no issues each of these 16 configurations being equally likely, they may struggle with what we call the *fundamental postulate of equilibrium statistical mechanics*, or the *postulate of equal a priori probabilities* later on in this course. This fundamental postulate is essentially no different from us assuming equal probability for each of the 16 configurations here. So if you find yourself struggling with this point then, come back here and remind yourself that is really isn't different from this!

Fluctuations

For small $N = 4$, a fluctuation where all 4 molecules are one the left side of the box would occur moderately frequently – about 1 in 16 frames, as computed previously. If we were to plot the relative number of particles on the left side of the box as a function of the frame index, we would obtain a graph like below, for instance.
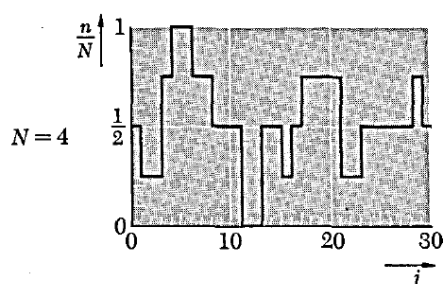

Figure 1.2: Relative number n/N of particles in the left hand half of box as time passes for $N = 4$

1.2.2 Larger Systems

Probability of *one* configuration/situation where $n = N$

As N increases however, we note that the probability of all N molecules being on the left becomes negligibly small, because there's always only one configuration that achieves this.

$$P_N = P_0 = \frac{1}{2^N}$$

Distributions of molecules that are *nonuniform* such that $n \ll N/2$ or $n \gg N/2$ correspond to relatively few configurations. Situations of these kinds that can be obtained in relatively few ways are said to be *nonrandom* or *orderly*[b].

The number of configurations for $n \approx n' \approx N/2$ however increases with N, and the larger the N, the more configurations will correspond to this particular situation. Situations of these kinds that can be obtained in many different ways is said to be *random* or *disordered.* We can (and will) further quantify these statements in the next chapter, but for now an instinctive appreciation of the matter will do.

> ***Random/disordered***: A situation that can be achieved in many ways, i.e. a situation that has many configuration corresponding to the particular macroscopic description. (In our illustration the macroscopic description is the number of molecules on the left side of the container, n)

Box 1.2: Putting things into Context

Let us put these things into context. When $N = 4$, we expect 1 in 16 frames in our movie strip to have all 4 molecules on the left. With $N = 80$, we would have $2^{80} \approx 10^{24}$ possible configurations. We would expect 1 in 10^{24} frames to have all 80 molecules on the left. If we took a million pictures every second, you would have to run the film for a time appreciably greater than the age of the universe before we have a reasonable chance of obtaining one frame showing all molecules in the left half of the box[a]. Consider now a realistic example of a box having volume 1cm^3, containing air at atmospheric pressure and room temperature. It would have $N = 2.5 \times 10^{19}$ molecules, which would lead us to approximately $10^{2.5 \times 10^{18}}$.

---
[a]For context, there are 3.15×10^7 seconds in a year, and the estimated age of the universe is of the order of 10^{10} years.

Probability of a *multiple* configurations/situation of a *range* of possible ns

It is therefore clear that when N is large, large fluctuations are exceedingly rare. However, one might argue that small fluctuations would be similarly common, and that if we group the $N = 50$ into 5 ranges, the probabilities wouldn't differ too significantly the Table 1.1 above. But upon doing the computation (done in tutorials), we note that the actual probabilities work out to be:

No of particles on left, n	$0 \leq n \leq 10$	$10 < n \leq 20$	$20 < n < 30$	$30 \leq n < 40$	$40 \leq n \leq 50$
Probability, P_n	1.19×10^{-5}	0.101	0.797	0.101	1.19×10^{-5}

Table 1.2: Probabilities for 4 different macroscopic scenarios when $N = 50$

These probability values are solely due to the fact that there are more ways to achieve (that is, more configurations that correspond to) the more *random* situation (i.e where n does not vary appreciably from 25).

Fluctuations and Equilibrium

From Table 1.2 above, we can see that for a system where the total number of particles is larger, fluctuations corresponding to an appreciably nonuniform distribution of particles (i.e. $0 \leq n < 10$ in our case, for instance) occurs

---
[b]I find Reif's choice of words here rather confusing. Molecular distribution that are *not uniform* are seen as *not random* or *orderly*. I am sticking with this choice of words to follow Reif, but do take note as it is easy to be confused here.

almost never! Plotting a similar graph of the relative number of molecules on the left hand side of the box may (for instance) result in,
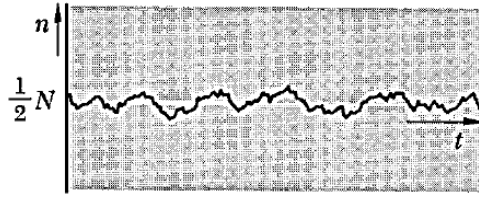

Figure 1.3: Relative number n/N of particles in the left hand half of the box as time passes for large N

A macroscopic description, or macrostate, which does not tend to change much in time is said to be in *equilibrium*. That is, if the description of the macrostate is such that the fluctuations are captured within that description in a way that the macrostate ends up not changing much in time, this macrostate is said to be in equilibrium. For instance, the macrostate where there are 20-30 molecules (out of 50) on the left side of the box is a state that does not change much in time. A system initially with 20-30 molecules on the left of the box will remain in that state a whooping 80% of the time. Whilst this 80% may not be sufficiently high to qualify as equilibrium, one can imagine that a state with 200-300 molecules (out of 500) which has a $P_{200 \leq n < 300} = 0.999991$, would definitely qualify as an equilibrium state. A system initially in that macrostate will remain in that state 99.991% of the time! Recall further that in a realistic situation, N is of the order of 10^{19}!

> ***Equilibrium***: A macroscopic state which does not tend to change in time

1.3 Approach to Equilibrium & Irreversibility

Approach to Equilibrium

For the same system with $N = 50$, consider a scenario where $0 \leq n < 10$ at a particular time τ. It matters not how we got to this highly improbable macrostate, all that matters is that the system now is in this state. What can we now say about the probability of each macrostate in the next time step? Clearly Table 1.2 no longer holds as the each configuration is no longer equally probable given the initial condition of this system.

In order to preserve this macrostate however, the particles must all move in a particular orderly way. Although we may not want to numerically compute the probability of this, it would align with our understanding that the probability of all the particles moving in such an orderly way would be low. Note further that to *continuously* preserve that macrostate over a course of time is even more unlikely!

As time passes the random motion of the particles will inevitably lead to the situation where this random motion no longer has a discernible effect on the system. By definition, this would be our equilibrium state, since this means the macrostate no longer changes in time. This situation would be the scenario where $n \approx N/2$ (see Fig. 1.4 below). Note that what's at the heart of this behaviour is nothing other than what is known to be the second law of thermodynamics!

> ***Second law of Thermodynamics (Primitive version)***: If an *isolated* system is in an appreciably non-random situation, it will generally evolve in time so as to approach ultimately it's most random situation where it is in equilibrium.

This is a very primitive version of the second law of thermodynamics. We will sharpen our notion of this second law repeatedly over the course of these lectures to reach our final form, and we will further discuss the other laws in turn.

Irreversibility

Our discussion above demonstrates that when an isolated macroscopic system changes in time, it tends to do so in a very definite direction – from a less random to a more random situation. Now, let the *time-reversed* process be the
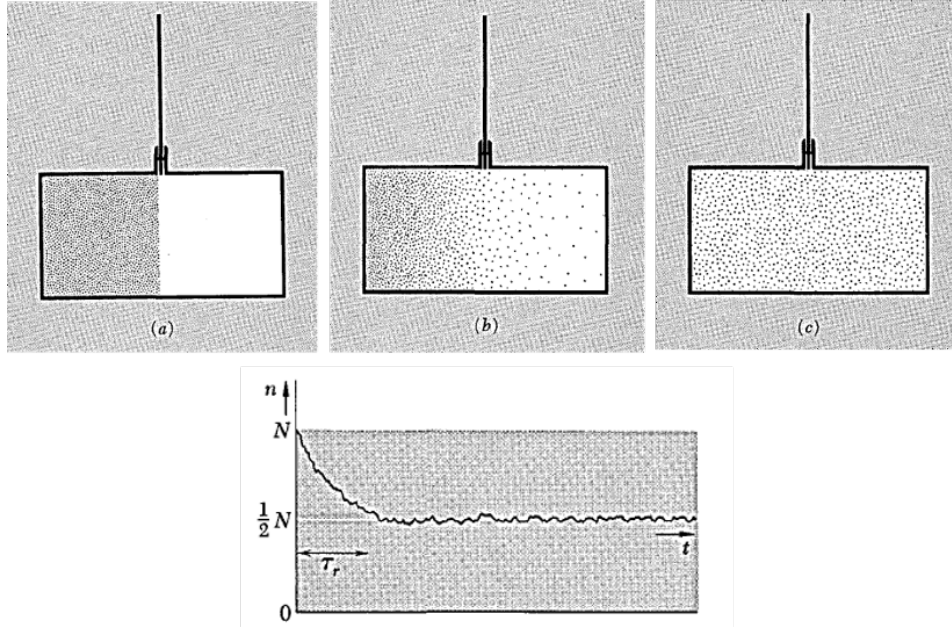
Figure 1.4: Example scenario where initially all n molecules are on the left half of the container due to some external intervention. Probabilistic arguments applied to the velocity of the molecules now dictate that the above course of event is bound to happen. Bottom graph depicts the relative number n/N of particles in the left hand half of the box as time passes given this initial condition.

process that would occur if one imagined the direction of time to be reversed[c]. With this conception of a time-reversed process, we are ready to briefly describe an irreversible process.

> **Irreversible process**: An *irreversible* process is a process where the time-reversed version of the process would almost never occur in reality.

It should be clear that there is nothing intrinsic in the law of motion of the particles of a system which gives rise to a preferred direction of time. This preferred direction arises only when one deals with an *isolated* macroscopic system which is somehow *known* to be in a special nonrandom situation initially. Take for instance the scenario where a system arrived at it's orderly state of $n \gg N/2$ due to a very rare spontaneous large fluctuation (left figure below). The forward-time and reversed-time process of that scenario would not be distinguishable unlike the scenario on the right where a specific direction of time is singled out by the knowledge that there was an external interaction prior to it being left undisturbed, as discussed above.
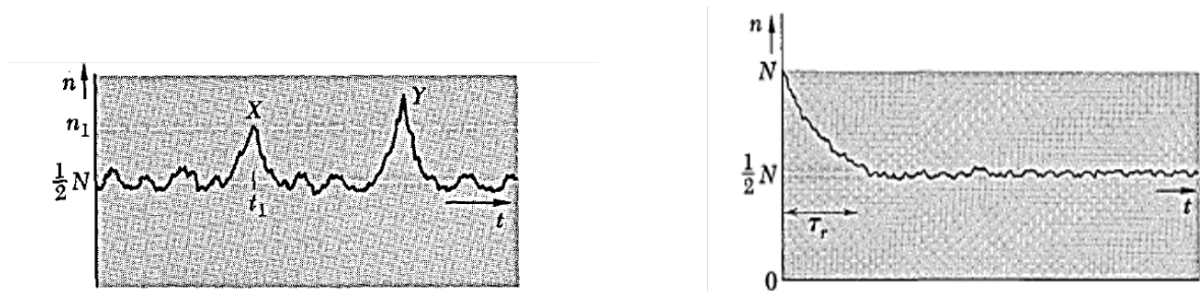
Figure 1.5: Left: No preferred direction of time for the scenario where a system left undisturbed got into a special orderly distribution of $n \gg N/2$ as a result of a very rare spontaneous fluctuation. Right: Preferred direction singled out by knowledge that before being left undisturbed, it was made to interact with some external system.

---

[c]If we consider taking a video of the system, the time-reversed process would be the video played backwards

Figure 1.6: In the world we're familiar with, we would never expect an *isolated* system to proceed in the time-reversed direction.

Note further that there is a *degree* to the irreversibility in that it becomes more pronounced the more particles there are as the occurrence of an orderly situation is increasingly unlikely with the larger number of particles. The concept of irreversibility would not mean much in the scenario where $N = 4$.

1.4 The Equilibrium Scenario

Just to summarize, in the event it wasn't already clear, a system in equilibrium is particularly simple as the following hold true:

- The macrostate of a system in equilibrium is time-independent, in the sense that the average value of the macroscopic parameter remains constant in time.

- The macrostate of a system in equilibrium is the most random macrostate under the specified condition

- The above point implies that the equilibrium macrostate of the system is independent of its past history (see our discussion in the previous section. It does not matter whether the system originally had a rare large fluctuation or a prior intervention. The final equilibrium state of both scenarios would not be distinguishable).

- We will learn that precisely because the equilibrium macrostate is the most random macrostate, it can be completely specified by very few macroscopic parameters.

Chapter 2

Probability Basics

Overview:
In this chapter we develop a thorough understanding of statistical ensembles and its relationship to the notion of probabilities. We further discuss some elementary relations that probabilities obey and explore some descriptive statistical measures that will allow us to characterize distributions without explicitly specifying the full probability distribution. We will end off the chapter by applying these tools in a sample two-outcome system.

Intended Learning Outcomes:
At the end of this chapter, students should be able to:

1. *explain* the statistical ensemble and its utility.
2. *explain* the concept of equilibrium by considering the statistical ensemble
3. *derive* the elementary relations of probabilities that we utilize from the basic notion/definition of probability.
4. *explain* the significance of two statistical measures we will utilize to describe the distribution, namely the mean and the dispersion/spread (standard deviation) with reference to the statistical ensemble
5. *illustrate* the application of said tools in a sample system with two discrete outcomes.

Textbook References:
Reif: 2.1-2.5

Flow of chapter:

1. Statistical Ensemble
2. Elementary Relations among Probabilities
3. Descriptive Statistical Measures
4. Calculation of Mean Values
   - Illustration II
   - General results
   - Explicit results

Probabilities are indispensable in daily life. We utilize it in games of chance, in insurance business and sampling procedures for polls. They play a fundamental in quantum mechanics and most importantly for us, they will form the basis of our entire discussion of macroscopic systems.

2.1 Statistical Ensembles

Consider a system A, on which we can perform experiments. Often times, the outcome of one single experiment cannot be predicted with certainty for one of two reasons:

- It is intrinsically impossible
- We do not possess sufficient information to permit such a unique prediction

This leads us to a statistical description of our experimental outcomes. Instead of one single system of interest, we now consider an *ensemble* consisting of a very large number $\mathcal{N}$ of "similar" systems[a]. By similar, we mean each system satisfies the same conditions known to be satisfied by system A. We then enumerate all possible mutually exclusive outcomes of the experiment, and ask: in what fraction of cases would a particular outcome of an experiment occur. This fraction

$$P_r \equiv \frac{\mathcal{N}_r}{\mathcal{N}} \tag{2.1}$$

is called the probability of occurrence of outcome r. Therefore, although we cannot know the outcome of a single experiment, we now predict the probability of occurrence of a particular outcome.

Box 2.1: The role of information in statistical ensembles

Probability theory is a lot more complex and nuanced that most realize. The interpretation of probability is in fact a matter of serious debate in some circles. There are two main camps (of which, there may be sub-camps): the frequentist interpretation and the Bayesian interpretation. The frequentist interprets probabilities as the frequency of particular outcomes under repeated trials (objective), whereas the Bayesians see probability as a degree of *belief* (subjective).

Most scientific presentation (this included) presents a frequentist version of probabilities. Some do include the nuances of the Bayesian interpretation (this presentation included), by alluding to the information necessary to construct the ensemble to begin with. The probability of the occurrence of any outcome depends crucially on the information available about the system under consideration. This information will affect the *statistical ensemble* chosen, which will then affect the fraction that is our probability.

The example of Reif captures it well. If we're interested in the probability that a person will be hospitalized at some time between the ages of 28-32, our statistical ensemble will be a large number of people of that age group. However, if we were told that the person is female, our ensemble will now be a large number of *women* of that age group. The resulting fraction may differ significantly as women of that age group may be hospitalized due to reasons men may be exempt from – childbirth, for instance.

When considering a statistical ensemble of systems evolving in time, do note the distinction between

- The behaviour of a single system, say system k as a function of time
- Probabilistic statements about the system at any one time t

Your Illustration: Statistical Ensemble

---

[a]In principle, $\mathcal{N}$ is imagined to be arbitrarily large in the sense of $\mathcal{N} \to \infty$

With this definition of a statistical ensemble, we are ready to redefine our concept of equilibrium. A statistical ensemble of systems is said to be *time-independent* if the number of its systems exhibiting any particular outcome is the same for every time step. Previously, without the definition of the statistical ensemble, the fluctuation in any one system prohibited us from discussing the outcomes. Although the outcome of any one single system is not well-defined due to fluctuations, the *fraction* of systems exhibiting this outcome over the ensemble always has a well defined value.

> **Equilibrium**: An isolated macroscopic system is said to be in equilibrium if a statistical ensemble of such systems in time-independent.

Note that in the previous chapter, we said that "A system initially with 20-30 molecules on the left side of the box will remain in that state a whooping 80% of the time." We then proceeded to comment that that percentage may not be sufficiently high to qualify as equilibrium. At that juncture, our language was not sufficiently precise to determine if that was scenario was indeed an equilibrium scenario. Now however, we can. Consider the outcome of 20-30 particles being found on the left. If the fraction of systems displaying that an outcome in one particular time-step is equivalent to the fraction of system displaying that outcome in the next time-step, that would be the equilibrium state!

Box 2.2: Ensemble average versus time average

Note that when the system is in equilibrium, then the ensemble of such a system is *time-independent*. In such a scenario, a large number of *successive* observations on a *single* system will then be equivalent to a large number of *simultaneous* observations on *many* systems. That is the ensemble average at a specific time (averaging over the rows down one single column) is indistinguishable from the time average (averaging over the columns across one single row). This concept of the ensemble average being indistinguishable from the time average, is sometimes known as ergodicity.

2.2   Elementary Relations among Probabilities

These relations we are about to derive are almost self-evident. However going through the process with the concept of the statistical ensemble at the back of your mind would be beneficial.

Set-up

Suppose experiments on system A can lead to any of α mutually exclusive outcomes or events.

- We label each outcome or event by some index $r = 1, 2, 3 \ldots \alpha$
- In an ensemble of similar systems, $\mathcal{N}_1$ of them will exhibit outcome 1, $\mathcal{N}_2$ of them will exhibit outcome 2, ... and $\mathcal{N}_\alpha$ of them will exhibit outcome α.

Normalization

Probability of either one of two (or more) outcomes

Joint Probabilities

Consider now a scenario where we do two different experiments of the system. We

- label the α possible mutually exclusive outcomes of the first experiment with the index $r = 1, 2, 3 \ldots, \alpha$,
- label the β possible mutually exclusive outcomes of the second experiment with the index $s = 1, 2, 3 \ldots, \beta$.
- denote $\mathcal{N}_{rs}$ to be the number of systems in an ensemble of $\mathcal{N}$ systems that are characterized by the *joint* occurrence of *both* event r *and* event s.

In summary:

$$\text{Normalization:} \qquad \sum_{r=1}^{\alpha} P_r = 1 \qquad (2.2)$$

$$\text{Probability of mutually exclusive outcomes } i \text{ or } j: \quad P(r = i \text{ or } r = j) = P_i + P_j \quad (2.3)$$

$$\text{Joint probability distribution } (r, s \text{ statistically independent}): \qquad P_{rs} = P_r P_s \quad (2.4)$$

2.3 Descriptive Statistical Measures

In the previous section, we did not care about the value of the outcome, only how frequently it occurred. The reason for this was merely because we were sharpening out concept of probability. In reality, we generally do care about the value of the outcome.

Therefore, consider now a random variable u that can assume α possible distinct values

$$u_1, u_2, u_3, \ldots u_\alpha$$

with probability

$$P_1, P_2, P_3, \ldots P_\alpha,$$

where we know

$$P_r = \frac{\mathcal{N}_r}{\mathcal{N}}.$$

The most complete statistical description of the random variable u is given by the specifications of probability distribution P_r for all u_r values. However, one can also characterize the distribution of this random variable via other statistical measures. We will discuss the two most fundamental ones in this section.

2.3.1 Mean Value (i.e. Ensemble Average)

We are given a variable u that assumes possible values u_r with respective probability P_r. The mean $\bar{u}$ indicates the central value of u about which the various values of u_r are distributed, while taking the frequency of its occurrence into account.

Definition

Property I: Ensemble average of function of u, $\bar{f}(u)$

Property II: Ensemble average of sum of functions of u, $\overline{f(u) + g(u)}$

Property III: Ensemble average of a constant multiple of a function of u, $\overline{cf(u)}$

Property IV: Ensemble average of the product of the function of two statistically independent variables, $\overline{f(u)g(v)}$

Consider now two different variables u and v, which can assume the following values

$$u_1, u_2, u_3 \ldots u_\alpha$$

$$v_1, v_2, v_3 \ldots v_\beta$$

In summary: The central value of u about which the various values of u_r are distribution is but the ensemble average value of u, and given by,

$$\bar{u} = \sum_{r=1}^{\alpha} P_r u_r \tag{2.5}$$

The following is true about this average:

$$\bar{f}(u) = \sum_{r=1}^{\alpha} P_r f(u_r) \tag{2.6}$$

$$\overline{f(u) + g(u)} = \bar{f}(u) + \bar{g}(u) \tag{2.7}$$

$$\overline{cf} = c\bar{f}(u) \tag{2.8}$$

$$\overline{f(u)g(v)} = \bar{f}(u)\bar{g}(v) \tag{2.9}$$

Note that Eq. (2.9) only holds in the special case where u and v are statistically independent.

2.3.2 Dispersion

The mean value $\bar{u}$ returns us the central value of u about which the values of u_r are scattered about. However, we don't know anything about *how spread out* this distribution of u_r are. The dispersion is one measure to characterize this spread.

In summary: To measure the spread of u_r about $\bar{u}$, we consider the standard deviation

$$\Delta u = \left[\overline{(\Delta u)^2} \right]^{1/2} = \left[\sum_{r=1}^{\alpha} P_r \left(u - \bar{u} \right)^2 \right]^{1/2} \tag{2.10}$$

2.4 Calculation of Mean Values: Spin System Example

A knowledge of the complete probabilities P_r for all values of u_r will give us complete statistical information about the distribution of the values of u_r. However computing the actual values of P_r can prove difficult and time-consuming. Even the simplest case of a two-outcome N particle ideal system is arduous to do by hand (as will be seen in tutorials!).

In this section, we would like to see if any conclusions can be made about the mean values and the spread of these values, without actually specifying the complete underlying probability distribution of the macroscopic system. In particular, we are interested in the following two points:

- Could we link the mean value of the macroscopic parameter to the mean value of the corresponding microscopic property of a singular microscopic system? Could we do the same for the dispersion?
- Could we therefore then determine the mean value of the macroscopic parameter and its associated spread by just specifying the probability distribution of a singular microscopic system?

2.4.1 Set-up

> **Illustration II**
>
> Consider an ideal system[a] of N identical (assumed distinguishable) spin-$\frac{1}{2}$, each having associated magnetic moment μ_0.
>
> For the sake of generality, suppose the spin system is located in an external magnetic field **B**. The measurement of the component of the magnetic moment of the spin along a particular direction[b] can take only two possible values $+\mu_0$ (parallel to **B**) or $-\mu_0$ (anti-parallel to **B**). For simplicity, we will denote the first outcome as spin up, the second as spin down.
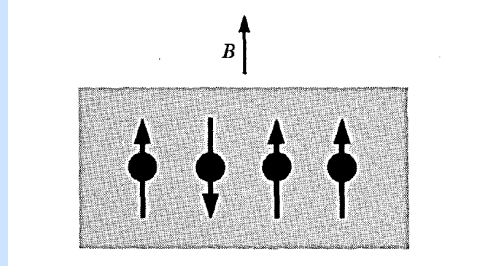>
>
>
> Figure 2.1: A system consisting of $N = 4$ ideal spins $\frac{1}{2}$. Each arrow indicates the direction of the magnetic moment of the spin. The direction of the external magnetic field is denoted by **B**.
>
> We denote the probability that a single spin points in the up direction as p, and the probability it points in the down direction as q, such that $p + q = 1$. When **B**$= 0$[c], there is no preferred direction in space, and therefore
>
> $$p = q = \frac{1}{2}.$$
>
> Let n be the number of magnetic moments pointing up and n' be the number of moments pointing down such that
>
> $$n + n' = N$$
>
> The total magnetic moment M is given by the algebraic sum of the individual magnetic moments over all the N particles
>
> $$M = \mu_1 + \mu_2 \ldots + \mu_N$$
>
> ---
>
> [a]The particles might be electrons, atoms with one unpaired electron, or nuclei such a protons.
> [b]Taken to be the direction of **B** here
> [c]In the presence of non-zero **B**, the magnetic moment will be more likely to point in the direction of the field than against it, $p > q$. For the moment, we will assume the value of p is given. We will learn to compute this value in a later chapter.

2.4.2 Relative Magnitude of Fluctuation

The total magnetic moment M of the system above is the outcome of the experiment that whose distribution we are concerned with. Therefore, in the following, we discuss the mean and dispersion of M where,

$$M = \sum_i^N \mu_i \tag{2.11}$$

Be well aware that the sum over i is over the *particles*, and not over the *outcomes* as the sum over r in the previous sections were.

Mean

Dispersion

$$\text{Mean}: \qquad \overline{M} = N\bar{\mu} \qquad\qquad (2.12)$$

$$\text{Dispersion/variance}: \qquad \overline{(\Delta M)^2} = N\overline{(\Delta\mu)^2} \qquad\qquad (2.13)$$

$$\text{Standard Deviation}: \qquad \underset{\sim}{\Delta} M = \sqrt{N}\,\underset{\sim}{\Delta} u \qquad\qquad (2.14)$$

[a]Note that although our particular system is a spin-$\frac{1}{2}$ system, the discussion in this subsection so far has not referred explicitly to the underlying probability distribution not the necessity for a two-outcome scenario. Therefore this discussion holds for any generic spin system

Discussion

In particular, note from Eq. (2.12) that the mean of M is proportional to N, the number of particles available, whereas from Eq. (2.14) the standard deviation of M is proportional to $\sqrt{N}$. Therefore the *relative magnitude* of $\underset{\sim}{\Delta} M$ compared to $\overline{M}$ *decreases* as N increases:

$$\frac{\underset{\sim}{\Delta} M}{\overline{M}} = \frac{1}{\sqrt{N}}\frac{\underset{\sim}{\Delta}\mu}{\bar{\mu}} \qquad\qquad (2.15)$$

It is hard to fully appreciate the importance of this result right now, but it is crucial to the understanding of the second law down the line. This is basically at the heart of the argument that when N is large, the probability distribution becomes very sharply peaked – because the spread of the distribution, relative to the mean of the distribution, decreases. Note that the spread itself *increases* as can be seen in Eq. (2.14), but the spread *relative* to the mean decreases. Recall further that our course deals with insanely large values of N.
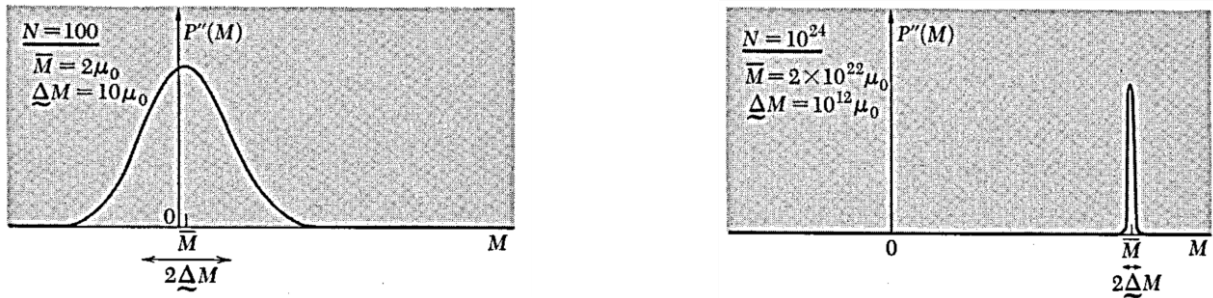

Figure 2.2: Example taken from the textbook that demonstrates how the probability distribution becomes very sharply peaked when $N \sim 10^{24}$

We will come back to this argument many times over the course of these lectures. This is also related to the reason why the values in Table 1.2 are the way they are. Compared to 4 particles, the probability distribution of 50 particles is a lot more sharply peaked! We can (and will) work out those numbers in tutorials, but the numbers alone may not be able to convey the deeper meaning associated with it.

> ***Relative magnitude of fluctuation***: The relative magnitude of the <u>standard deviation</u> compared to the <u>mean</u> of a quantity is a good measure of the *relative magnitude of fluctuation* of that quantity. We will find that for most if not all quantities, this ratio scales as per $1/\sqrt{N}$, decreasing with increasing N. In fact, at the values of N that we consider for statistical mechanics ($N \sim 10^{24}$), this relative magnitude of fluctuations are so small that our probability distributions are sharply peaked about the mean value.

2.4.3 Explicit Results for Spin$-\frac{1}{2}$ System

Our particular spin-$\frac{1}{2}$ system is a two-outcome system. Keep in mind the statistical ensemble during this discussion, as it is extremely easy to confuse between the average across the ensemble and the sum across the particles.

Your Illustration:

Our aim in this section is to obtain an explicit expression for $\bar{\mu}$ and $\underset{\displaystyle\sim}{\Delta}u$ for our system of N ideal spin-$\frac{1}{2}$ to insert into Eq. (2.12) and Eq. (2.14). To that end, recall that each spin μ can take only two values:

- μ_0 with probability p, or
- $-\mu_0$ with probability $q = 1 - p$.

In summary: For our spin$-\frac{1}{2}$ system,

$$\text{Mean}: \qquad \overline{M} = N(p-q)\mu_0 \qquad (2.16)$$

$$\text{Dispersion/variance}: \qquad \overline{(\Delta M)^2} = 4Npq\,\mu_0^2 \qquad (2.17)$$

$$\text{Standard Deviation}: \qquad \underset{\sim}{\Delta} M = 2\sqrt{Npq}\,\mu_0 \qquad (2.18)$$

2.4.4   Discussion

We started out this section by introducing the system we would discuss. We then proceeded to link the mean of the macroscopic parameter (total magnetic moment) to mean of the property (magnetic moment) of an individual spin particle. We did the same for the dispersion. Lastly, we obtained the expression of the mean and associated dispersion of the macroscopic parameter by specifying the full probability distribution of the property for the individual spin particle.

Note that the full probability distribution of the *macroscopic* system would be far more difficult to specify! We will see this in our tutorials. As N increases, specifying the full probability distribution of the macroscopic system quickly gets out of hand. However, our discussions here allow us to compute the mean and spread of the macroscopic parameter without requiring the full probability distribution!