MATH327: More Is Different
Statistical Mechanics, Thermodynamics, and All That

David Schaich **Spring Term 2026**

LECTURE NOTES

Last modified 19 January 2026

Module information and logistics 2

Unit 1: Central limit theorem and diffusion 8

Unit 2: Micro-canonical ensemble 26

Unit 3: Canonical ensemble 40

Unit 4: Ideal gases 56

Unit 5: Thermodynamic cycles 68

Unit 6: Grand-canonical ensemble 80

Unit 7: Quantum statistics 90

Unit 8: Quantum gases 101

Unit 9: Interacting systems 127

Unit 10: Synthesis and broader applications 151

Module information and logistics

Coordinator

David Schaich
Theoretical Physics Wing Room 123
david.schaich@liverpool.ac.uk
calendly.com/daschaich
www.davidschaich.net

Overview

"More Is Different" is the title of a famous 1972 essay by Philip Anderson, which established the concept of emergent phenomena — the idea that large, complex physical systems generally can't be understood by extrapolating the properties of small, simple systems. For example, consider the $\sim 10^{22}$ H_2O molecules in a cubic centimetre of water. While we can use Newton's laws (or the laws of quantum mechanics) to analyse a few of these molecules, this does not allow us to predict processes such as phase transitions of this water into steam or ice.

Instead, we have to apply the stochastic (i.e., probabilistic) techniques of *statistical mechanics* — one of the central pillars of modern physics, along with quantum mechanics and relativity. While statistical mechanics was originally developed in the context of thermodynamics in the nineteenth century, it is more generally applicable to any large-scale (macroscopic) behaviour that emerges from the microscopic dynamics of *many* underlying objects. It is intimately connected to quantum field theory, and has been applied to topics from nuclear physics and cosmology to climate science and biophysics, often with outstanding success (recently recognized by the 2021 and 2024 Nobel Prizes in Physics, to name just two).

The module outline on the previous page is organized around the concept of *statistical ensembles* introduced in the early 1900s. In essence, a statistical ensemble is a mathematical framework for concisely describing the properties of idealized physical systems subject to certain constraints. After studying the probability foundations underlying these frameworks, we meet the *micro-canonical ensemble* in unit 2 and the *canonical ensemble* in unit 3. The following two units 4–5 apply the canonical ensemble to investigate non-interacting ("ideal") gases and thermodynamic cycles. Unit 6 introduces a third statistical ensemble, the *grand-canonical ensemble*, which units 7–8 apply to several types of non-interacting quantum gases. (No prior exposure to quantum mechanics is required — see below for more information.) Finally, in unit 9 we begin to explore the effects of interactions, which open up a much broader landscape of applications that we will survey for the remainder of the term.

Schedule

Most weeks we will have the following schedule:

- Lecture at 10:00–11:00 on Tuesdays in Elec. Eng. Lecture Room E5 (205)
- Tutorial at 10:00–11:00 on Wednesdays in Room 106
- Lecture at 11:00–13:00 on Thursdays in Room 106

The tutorials in weeks 2, 3 and 4 (on 4, 11 and 18 February) will be computer lab sessions in Hub 502 PC Teaching Centre B, to provide opportunities for you to work on the computer assignment summarized below.

I will use Panopto to record lectures (and lecturey bits of tutorials and computer labs). These recordings will appear along with all other resources on our Canvas site,
canvas.liverpool.ac.uk/courses/84431
While these recordings will allow you to catch up if you must miss a lecture, you will gain the most benefit from attending in person so that you can ask questions and check your understanding.

Office hours will take place at 11:00–12:00 on Tuesdays and Wednesdays, following the corresponding class meeting. They will be hybrid, taking place in Room 123 of the Theoretical Physics Wing, with the option of joining via Zoom at this link (meeting ID 819 4732 3591, passcode Math327!).

If these times do not work with your schedule, you can also make an appointment through calendly.com/daschaich, use the Canvas discussion board (where anonymous posting is enabled), or reach me by email at david.schaich@liverpool.ac.uk. I will aim to respond to emails and discussion board queries within 48 hours.

Assessment and academic integrity

There will be three in-term assignments. Each accounts for 15% of the module mark, with the remaining 55% coming from the final exam. The deadlines listed below have been centrally coordinated within the Department to minimize pile-up across different modules.

15% A computer assignment due **Wednesday, 18 February**

30% Two equally weighted homework assignments, the first due **Friday, 6 March** and the second due **Friday, 24 April**

55% A two-hour in-person final examination to be centrally scheduled within the May exam period

According to the University's Code of Practice on Assessment (CoPA), Submissions more than 120 hours late will be awarded zero marks, though I will still endeavour to provide feedback on them. I will aim to return feedback and share model solutions within two or three weeks of the deadline for the homeworks or computer assignment, respectively.

For your other modules you already should have read and understood the Department's current academic integrity guidance as well as the Academic Integrity Policy detailed in CoPA Appendix L. If you have any questions about what is or is not acceptable, please ask me or our Academic Integrity Officer Alena Haddley. In all cases, the spirit of the Academic Integrity Policy should take precedence over legalistic convolutions of the text.

In particular, I encourage you to discuss the in-term assignments with each other, since discussing and debating your work is a very effective way to learn. Note that I say *your work* — your submissions for all assignments must be your own work representing your own understanding, and the examination must be done on your own. It is unacceptable to copy solutions in part or in whole from other students (current or prior) or from other sources (commercial or otherwise). Should you make use of resources beyond the module materials — including generative AI tools such as ChatGPT — these must be explicitly referenced in your submissions. Clear and neat presentations of your workings and the logic behind them will contribute to your mark.

Main resources and materials

The main materials we will use are the lecture notes you are currently reading. As you read further, you will encounter gaps in the notes, which provide bite-sized exercises to help you check your understanding. While we will fill most gaps during lectures, I encourage you to use them as opportunities to practice.

The ten units into which the content is organized won't neatly match up with the twelve weeks of the term. Some units will require more time than others. Regular Canvas announcements will summarize what we cover each week.

We will use 'natural units' in which the Boltzmann constant $k = 1$, and logarithms have base e unless otherwise specified (i.e., $\log x = \ln x$). There is no need to memorize any equations. Many equations are numbered so that they can be referenced later on, not necessarily because they are important. Key results, definitions and concepts are highlighted by coloured boxes, and you should aim to be confident in your understanding of these.

These lecture notes were first written 'live' during the 2021 and 2022 editions of this module. While they are now much more stable, they continue to be improved, refined and sometimes corrected. The "Last modified" date at the bottom of each page will flag any changes that occur during the term.

Expected background

No prior exposure to quantum mechanics or computer programming is required — all necessary information on these topics will be provided. I do anticipate that you have previously seen the standard deviation, the binomial coefficient

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

that counts the number of possible ways to choose k objects out of a set of $N \geq k$ total objects, and gaussian integrals,

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2}\, dx = \sqrt{\frac{\pi}{a}} \qquad\qquad a > 0.$$

Programming

You are welcome to complete the computer assignment using the programming language of your choice. I recommend Python, which is free, user-friendly, and very widely used around the world. During the first two weeks of the term we will review this demo that explains all the Python programming tools you'll need. Python is available on University computers and should work on personal computers. You can also write and run Python code using many cloud services, of which I have had the best experiences with Google's Colab and CoCalc.[1] You may need to create a free account, and you should make sure to save a local copy to reduce the risk of losing your work. Alternative languages could include C, Fortran, R, or even MATLAB (through the University's site license). I advise against using Maple, which may struggle to handle parts of the assignment.

How to get the most out of this module

At this point in your studies, this advice should be familiar, but it's worth repeating.

Come to class. This will ensure regular contact with the material, and help you check that you understand it. If the module is moving slower than you'd prefer, coming to class will give you opportunities to ask about more interesting extensions, applications or complications.

Before class, take a quick look at the upcoming pages in the lecture notes, and think about how any gaps could be filled. Look for the big ideas rather than digging in to every detail, and see if you have any questions (or objections) to raise in class. After class, take a closer look at the details, and make sure the gaps have been filled to your satisfaction. Even though the lecture notes reflect my plans for the module, they may not exactly match what happens in class, especially when questions arise. We may gloss over some topics that are explained clearly in the notes, and we may delve deeper into other topics that merit further consideration.

Work on the tutorial exercises and homework problems (including the computer assignment). The best way to learn mathematics is by doing mathematics. These exercises and assignments are designed to make you think and help develop your mathematical muscles. In particular, the tutorial exercises and homework problems will be harder than exam questions, since you'll have much more time to think about them — so take a look through them as soon as you have them, and don't leave them until the last minute. Afterwards, review the model solutions and feedback, to make sure any confusing points are resolved.

---

[1] I have had worse experiences with replit.com, onlinegdb.com, mybinder.org and trinket.io — use these at your own risk.

Ask questions. Ask questions you think you're supposed to know the answer to. Ask questions you think everyone else knows the answer to. (They don't.) Ask questions about the big ideas, the specific details, and the connections between them. The opportunity to ask questions is the main benefit of taking a module. You can ask me; you can ask your classmates; you can ask the additional resources below.

Additional resources

The optional additional resources listed below may be helpful. You can use the module Reading List on Canvas to reach our library's records for the books.

Resources at roughly the level of this module:

1. David Tong, *Lectures on Statistical Physics* (2012), www.damtp.cam.ac.uk/user/tong/statphys.html
2. MIT OpenCourseWare for undergraduate Statistical Physics I (2013) and Statistical Physics II (2005), ocw.mit.edu/courses/8-044-statistical-physics-i-spring-2013/ ocw.mit.edu/courses/8-08-statistical-physics-ii-spring-2005/
3. Daniel V. Schroeder, *An Introduction to Thermal Physics* (2021)
4. Harvey Gould and Jan Tobochnik, *Statistical and Thermal Physics with Computer Applications* (2021)
5. J. Allday and S. Hands, *Introduction to Entropy: The Way of the World* (2024)
6. C. Kittel and H. Kroemer, *Thermal Physics* (1980)
7. F. Reif, *Fundamentals of Statistical and Thermal Physics* (1965)

More advanced and more specialized resources, which may be useful to consult concerning specific questions or topics:

7. MIT OpenCourseWare for postgraduate Statistical Mechanics I (2013) and Statistical Mechanics II (2014), ocw.mit.edu/courses/8-333-statistical-mechanics-i-statistical-mechanics-of-particles-fall-2013/ ocw.mit.edu/courses/8-334-statistical-mechanics-ii-statistical-physics-of-fields-spring-2014/
8. R. K. Pathria and P. D. Beale, *Statistical Mechanics* (2021)
9. Sidney Redner, *A Guide to First-Passage Processes* (2001)
10. Pavel L. Krapivsky, Sidney Redner and Eli Ben-Naim, *A Kinetic View of Statistical Physics* (2010)
11. Kerson Huang, *Statistical Mechanics* (1987)
12. Andreas Wipf, *Statistical Approach to Quantum Field Theory* (2013)

13. Weinan E, Tiejun Li and Eric Vanden-Eijnden, *Applied Stochastic Analysis* (2019)
14. Michael Plischke & Birger Bergersen, *Equilibrium Statistical Physics* (2006)
15. Sacha Friedli and Yvan Velenik, *Statistical Mechanics of Lattice Systems* (2018)
16. L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part 1* (1969)

A general book about learning, emphasizing (among other things) the value of retrieval practice compared to re-reading lecture notes or re-watching videos:

17. Peter C. Brown, Henry L. Roediger III and Mark A. McDaniel, *Make it Stick: The Science of Successful Learning* (2014)
    A short summary video is also available

Programming resources:

18. MATH327 Python programming demo (2026)
19. Beginner's Guide to Python (2024)
20. W3Schools Python Tutorial (2024)
21. Software Carpentry tutorials:
    - Programming with Python (2024)
    - Plotting and Programming in Python (2024)
22. Stormy Attaway, *MATLAB: A Practical Introduction to Programming and Problem Solving* (2013)
23. B. Barnes and G. R. Fulford, *Mathematical Modelling with Case Studies: Using Maple and MATLAB* (2014)

In addition, there is a vast constellation of other online resources such as Stack Exchange and Wikipedia. These can be great places to *start* learning about a topic, but are often terrible places to *stop*.

Unit 1: Central limit theorem and diffusion

Introductory remarks: More Is Different

> *What is most exciting about our work is that it illuminates the chain of connections between, on the one hand, the simple underlying laws that govern the behavior of all matter in the universe and, on the other hand, the complex fabric that we see around us, exhibiting diversity, individuality, and evolution. The interplay between simplicity and complexity is the heart of our subject.* —Murray Gell-Mann (1969 Nobel laureate), 1996

Mathematical sciences such as physics aim to determine the laws of nature and understand how these govern experimental observations — both in everyday circumstances and under extreme conditions. This mathematical understanding is typically guided by reproducing a set of observations, with the resulting framework then used to make predictions for other “observables”.

Over the past few centuries this process has been tremendously successful, with theoretical physics accurately predicting experimental and observational results from sub-atomic through to cosmological scales. Modern physics labs can create a vacuum better than in outer space and the coldest temperatures in the known universe, as well as going to the other extreme to reach temperatures of millions of degrees and pressures millions of times atmospheric pressure at sea level. Amazingly, many aspects of these realms of physics can be described by mathematics developed centuries ago.[2]

A crucial aspect of this success is the *emergence* of complicated physical phenomena from simple underlying mathematical laws. This module’s name borrows from a famous 1972 essay by Philip Anderson (1977 Nobel laureate), “More Is Different”, which helped to establish modern perspectives on emergent phenomena. We will focus on one particular domain in which simple mathematical principles enable amazing predictive capabilities: statistical mechanics (sometimes called probabilistic mechanics, statistical physics or statistical thermodynamics). Initially developed in the context of thermodynamics in the nineteenth century, statistical mechanics remains a central pillar of modern physics, and will retain this position in years to come. The foundations of statistical mechanics lie in the use of probability theory to mathematically describe experimental observations and corresponding laws of nature that involve stochastic randomness rather than being perfectly predictable.

The lack of perfect predictability in statistical mechanics is a matter of practicality rather than one of principle. It arises due to working with a large number of degrees of freedom — that is, a large number of independent objects such as atoms. For illustration, Avogadro’s number $N_A \approx 6.022 \times 10^{23}$ is the large number of molecules in everyday amounts of familiar substances — about 18 grams of water or about 22 litres of air at sea-level atmospheric pressure. Specifying the positions and velocities of $\sim 10^{23}$ objects would require far more information than could be stored by even the biggest

---

[2] A famous 1960 essay by Eugene Wigner (1963 Nobel laureate), “The Unreasonable Effectiveness of Mathematics in the Natural Sciences”, and subsequent work in the philosophy of physics, elaborates on why this may be considered ‘amazing’. This module will not comment extensively on philosophy.

existing supercomputers. Statistical mechanics instead produces simple mathematical descriptions of large-scale properties such as temperature, pressure and diffusion, which are generally of such outstanding quality that the underlying ‘randomness’ is effectively invisible.

Historically, the difficulty of detecting the stochastic processes underlying such *thermodynamic* properties made it challenging to convince skeptics that atoms and molecules really exist. Ludwig Boltzmann, a prominent early developer of statistical mechanics, endured a constant struggle to defend his ideas, which likely contributed to his deteriorating mental health and eventual suicide in 1906. A significant step to convincingly establish the existence of atoms was Albert Einstein’s use of statistical mechanics to explain the observed “Brownian motion” of particles suspended in fluids — this work was part of Einstein’s “miracle year” in 1905, along with special relativity and early contributions to quantum physics. Jean Perrin soon verified Einstein’s predictions and used them to determine Avogadro’s number; he was awarded the 1926 Nobel Prize in Physics for helping to demonstrate “the discontinuous structure of matter”.

Applications of statistical mechanics continue to advance “our understanding of complex physical systems” — quoting the 2021 Nobel Prize shared between Syukuro Manabe, Klaus Hasselmann and Giorgio Parisi. Even more recently, the 2024 Nobel Prize was awarded to John Hopfield and Geoffrey Hinton for “using ideas from statistical physics” to lay the foundations for machine learning and artificial neural networks. Other modern topics we will encounter in this module include explaining why stars don’t collapse under the ‘weight’ of their own gravity, and identifying effects of dark matter in temperature fluctuations observable in the cosmic microwave background lingering from the early years of the universe.

In this unit we focus on some of the foundational mathematics that will underlie our later development and application of statistical mechanics. Looking back to Boltzmann’s times, we can consider the following question some of his critics might have asked: *If the pressure of a gas in a container results from molecules stochastically colliding with the walls of that container, then how can the pressure be so stable, rather than itself fluctuating stochastically?* The mathematical answer lies in the **law of large numbers** and the **central limit theorem**, which we will review and apply to the physics of diffusion in one dimension.

1.1 Probability foundations

We begin by building a more formal mathematical framework around the concept of probability, through a sequence of definitions. First, a random **experiment** $\mathcal{E}$ involves setting up, manipulating and/or observing some (physical or hypothetical) system that involves some element of randomness. Flipping a coin is a simple random experiment. For the statistical ensembles we will focus on in later units, a typical experiment will be to allow a collection of particles to evolve in time, subject to certain constraints.

Each time an experiment is performed, the world is observed to be in some **state** ω. The specification of the experiment and the state must include all objects of interest, and may include more besides. When flipping a coin, for example, the full state could contain information not only about the final orientation of the coin, but also about its position — where exactly did it land?

The **set of all states** Ω collects all possible states ω that the given experiment $\mathcal{E}$ can produce, and is therefore intricately tied to $\mathcal{E}$ itself.

We are generally not interested in all aspects of the full state ω. For example, we won't care where a flipped coin lands. Instead we're typically only interested in whether it lands heads up or tails up — and we may want to set aside any state that doesn't cleanly reflect those options. The **measurement** $X(\omega)$ extracts and quantifies this information, acting as a function that maps the state ω to a number that we can mathematically manipulate. If we repeat the given experiment $\mathcal{E}$ many times and carry out the measurement X on each resulting state ω_i, we will obtain a sequence of numbers $X(\omega_i)$ that behave as a *random variable*.

Acting with the measurement X on all of the possible states in the set Ω defines the **set of all outcomes** (or **outcome space**) A:

$$X : \Omega \to A.$$

That is, A collects all possible measurement results that the given experiment $\mathcal{E}$ and measurement X can produce. A can be finite, countably infinite, or uncountably infinite (i.e., continuous).

Let's consider some examples to clarify these definitions. With an experiment of rolling a six-sided die and measuring the number $(1–6)$ that comes out on top, what is the set of all outcomes A? What additional information could be included in a corresponding state ω?

What is the outcome space A if we toss a coin four times and measure whether it lands heads up (H) or tails up (T) each time?

What information could characterize a state ω for a gas of 10^{23} argon atoms in a container? What might be interesting to measure?

For convenience, we can introduce a unique number as a *label* to characterize each state ω in the set Ω. Generalizing the concept of measurement, this provides a label function $L(\omega)$ that also behaves as a random variable. Our condition of uniqueness makes $L(\omega)$ isomorphic, so that the label can be used interchangeably with the full state, $\omega \longleftrightarrow L(\omega)$.

While the measurements $X(\omega)$ we consider generally will not produce a unique number for each ω, we will design them precisely to remove irrelevant information that doesn't interest us. Ignoring that irrelevant information leaves us free to interchange the set of outcomes A for the set of states Ω. (Some textbooks may never distinguish between A vs Ω in the first place, though this can be a source of confusion.)

Only a couple of definitions remain. The next is to define an **event** to be any subset of the set of all outcomes A. For example, events resulting from rolling a die could include (i) rolling a 6, (ii) rolling anything but a 6, (iii) rolling any even number, and many more. Collecting all events of interest defines the **set of events** (or **event space**) $\mathcal{F}$.

We are now prepared for the final foundational definition in this section — the **probability** P of an event in the set $\mathcal{F}$. Mathematically, P is a *measure function*,

$$P : \mathcal{F} \to [0, 1],$$

which must satisfy the following two requirements:

1. The probability of a countable union of mutually exclusive events must equal the sum of the probabilities of each of these events.
2. The probability of the outcome space ($\mathcal{F} = A$) must equal 1, even if A is uncountable. This simply means that the experiment $\mathcal{E}$ must produce a measurable outcome. We discard any experiment that doesn't produce such an outcome.

Combining the outcome space, event space and probability measure gives us a *probability space* $(A, \mathcal{F}, P)$.

For example, consider an experiment that can only produce N possible states, so that

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}.$$

As described above, the subscript is a label, and it is possible for two different states $\omega_i \neq \omega_j$ to produce the same measurement outcome $X(\omega_i) = X(\omega_j)$. This means that the size n of the outcome space A may be smaller than the size of Ω, $n \leq N$. We can write

$$A = \{X_1, X_2, \cdots, X_n\},$$

where each X_α is distinct and its label does not necessarily match the one on ω_i. We can take the individual X_α themselves to be the events we're interested in, choosing the event space

$$\mathcal{F} = \{X_1, X_2, \cdots, X_n\} = A. \tag{1}$$

These events are all mutually exclusive, so if we assign them probabilities

$$P(X_\alpha) \equiv p_\alpha \qquad \text{for } \alpha = 1, \cdots, n,$$

then the above requirements on probabilities demand that for any $\alpha \neq \beta$ we have

$$P(X_\alpha \text{ or } X_\beta) = p_\alpha + p_\beta$$

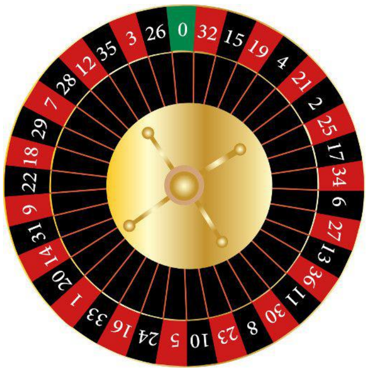
$$P(A) = P(X_1 \text{ or } X_2 \text{ or } \cdots \text{ or } X_n) = \sum_{\alpha=1}^{n} p_\alpha = 1.$$

Similarly choosing $\mathcal{F} = A$ for the six-sided die considered in an earlier gap, what are the probabilities p_1 through p_6 that result from assuming the die is fair?

Again taking $\mathcal{F} = A$ for the case of tossing a coin four times, what are the probabilities p_α that result from assuming the coin is fair? If we instead consider the event space

$$\mathcal{F} = \{\text{equal number of } H \text{ and } T, \text{ different numbers of } H \text{ and } T\},$$

what are the probabilities p_{equal} and p_{diff} for the two events in this $\mathcal{F}$?



The standard European roulette wheel shown to the left (source) has 37 pockets labelled “0” through “36”. 18 of these pockets are coloured red, 18 are coloured black and 1 (pocket “0”) is coloured green. Let an experiment be a spin of the roulette wheel, measuring the label of the pocket where the ball comes to rest (which also provides the pocket’s colour).

What is the outcome space A for a spin of the roulette wheel? With $\mathcal{F} = A$, what are the probabilities p_α for a fair wheel? With

$$\mathcal{F} = \{\text{ball in a red pocket, ball in a black pocket, ball in the green pocket}\},$$

what are the corresponding probabilities p_{red}, p_{black} and p_{green}?

The process of assigning probabilities to events is called *modelling*. The gaps above demonstrate that *symmetries* are a powerful way to constrain probabilities. The symmetry between the six sides of a fair die, the two sides of a fair coin, and the 37 pockets of a fair roulette wheel each sufficed to completely determine the corresponding probabilities p_α.

Modelling can also be guided by empirical data obtained by repeating an experiment many times. For example, if we don’t know whether a set of dice are fair, we will be able to infer their probabilities p_α (with a certain confidence level) by rolling them enough times. The need to repeat the experiment many times comes from the law of large numbers, to which we now turn.

1.2 Law of large numbers

Let’s return to the setup leading to Eq. 1 above, with

$$\mathcal{F} = A = \{X_1, X_2, \cdots, X_n\}$$

for finite n, and probabilities $p_\alpha = P(X_\alpha)$ that obey

$$p_\alpha \in [0,1] \qquad\qquad \sum_{\alpha=1}^{n} p_\alpha = 1.$$

We can generalize this notation by writing instead

$$\sum_{X\in A} P(X) = 1,$$

and introducing similar expressions for the **mean** μ and **variance** σ^2 of the probability space,

$$\mu = \langle X \rangle = \sum_{X\in A} X\, P(X) \tag{2}$$

$$\sigma^2 = \left\langle (X-\mu)^2 \right\rangle = \sum_{X\in A} (X-\mu)^2\, P(X). \tag{3}$$

The angle bracket notation indicates the **expected** (or **expectation**) **value** with general definition

$$\langle f(X) \rangle = \sum_{X\in A} f(X)\, P(X), \tag{4}$$

which is a linear operation,

$$\langle c \cdot f(X) + g(X) \rangle = c \, \langle f(X) \rangle + \langle g(X) \rangle \, .$$

The square root of the variance, $\sqrt{\sigma^2} = \sigma$, is the **standard deviation**. What is σ expressed in terms of $\langle X^2 \rangle$ and $\langle X \rangle^2$?

We now define a new experiment that consists of *repeating* the original experiment R times, with each repetition independent of all the others. Using the same measurement as before for each repetition, we obtain a new outcome space that we can call B. For $R = 4$, what are some representative outcomes in the set B? What is the total size of B?

Each outcome in B contains R different $X^{(r)} \in A$, one for each repetition $r = 1, \cdots, R$, and each with mean $\left\langle X^{(r)} \right\rangle = \mu$ and variance $\left\langle (X^{(r)} - \mu)^2 \right\rangle = \sigma^2$. Considering the case $R = 4$ for simplicity, any element of B can be written as $X_i^{(1)} X_j^{(2)} X_k^{(3)} X_l^{(4)} \in B$ with corresponding probability

$$P_B\left(X_i^{(1)} X_j^{(2)} X_k^{(3)} X_l^{(4)}\right) = P_A\left(X_i^{(1)}\right) P_A\left(X_j^{(2)}\right) P_A\left(X_k^{(3)}\right) P_A\left(X_l^{(4)}\right),$$

using subscripts to distinguish between the probability spaces for the single experiment (A) and repeated experiment (B).

Averaging over all R repetitions defines the *arithmetic mean*

$$\overline{X}_R = \frac{1}{R} \sum_{r=1}^{R} X^{(r)}. \tag{5}$$

Unlike the true mean μ, the arithmetic mean $\overline{X}_R$ is a random variable — a number that may be different for each element of B. That said, $\overline{X}_R$ and μ are certainly related, and so long as the standard deviation exists — that is, so long as σ^2 is finite — this relation can be proved rigorously in the limit $R \to \infty$.[3]

Here we will not be fully rigorous, and take it as given that

$$\left\langle \left(X^{(i)} - \mu\right)\left(X^{(j)} - \mu\right) \right\rangle = \sigma^2 \delta_{ij} = \begin{cases} \sigma^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases},$$

where the *Kronecker delta* $\delta_{ij} = 1$ for $i = j$ and vanishes for $i \neq j$. This is a consequence of our requirement that each repetition is independent of all the others. Using this result and the relation $\left(\sum_i a_i\right)\left(\sum_j b_j\right) = \sum_{i,j} (a_i b_j)$, express the following quantity in terms of σ and R:

$$\left\langle \left(\frac{1}{R} \sum_{r=1}^{R} X^{(r)} - \mu \right)^2 \right\rangle =$$

---

[3] In the computer project we will numerically investigate a situation where σ^2 diverges.

You should find that your result vanishes in the limit $R \to \infty$, so long as σ^2 is finite. Since the square makes this expectation value a sum of non-negative terms, it can vanish only if every one of those terms is individually zero.

This establishes the **law of large numbers**:

$$\lim_{R\to\infty} \frac{1}{R} \sum_{r=1}^{R} X^{(r)} = \mu, \tag{6}$$

where we have assumed finite $\left\langle X^{(r)} \right\rangle = \mu$ and $\left\langle (X^{(r)} - \mu)^2 \right\rangle = \sigma^2$.

1.3 Probability distributions

It is not necessary to make the assumption (Eq. 1) that our outcome space contains only a countable number of possible outcomes. The considerations above continue to hold even if the random variable X is a continuous real number. In this case, however, the identification of probabilities with outcomes is slightly more complicated, which will be relevant when we consider the central limit theorem in the next section.

When the outcome can be any number on the real line, the fundamental object is a **probability distribution** (or **density function**) $p(x)$ defined for all $x \in \mathbb{R}$. Starting from this density, a probability is determined by integrating over a given interval. Calling this interval $[a, b]$, the integration produces the probability that the outcome X lies within the interval,

$$P\left(a \le X \le b\right) = \int_a^b p(x)\, dx\,.$$

We similarly generalize the definition of an expectation value (Eq. 4) to an integral over the entire domain of the probability distribution,

$$\langle f(x) \rangle = \int f(x)\; p(x)\, dx\,.$$

We will omit the limits on integrals over the entire domain, so for $x \in \mathbb{R}$ we implicitly have $\int dx = \int_{-\infty}^{\infty} dx$, with $\int p(x)\, dx = 1$. An important set of expectation values is

$$\left\langle x^{\ell} \right\rangle = \int x^{\ell}\; p(x)\, dx, \tag{7}$$

which provides the mean and variance of the probability distribution $p(x)$, through generalizations of Eqs. 2–3:

$$\mu = \langle x \rangle = \int x\; p(x)\, dx \qquad\qquad \sigma^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2\,. \tag{8}$$

The expression for the variance should be familiar from your determination of the standard deviation in an earlier gap. Unless stated otherwise, we will assume the mean and variance are both finite for the probability distributions we consider.

1.4 Central limit theorem

The central limit theorem is a central result of probability theory (hence its name). Over the years it has been expressed in several equivalent ways, and there are also many distinct variants of the theorem accommodating different conditions and assumptions. Here we are interested in applying rather than proving the central limit theorem; you can find proofs in many textbooks.

The version of the theorem we use in this module assumes we have N independent random variables $x_1, \cdots, x_N$, each of which has the same (finite) mean μ and variance σ^2. Such random variables are said to be *identically distributed*, and a common way to obtain them is to repeat an experiment N times, as we considered in Section 1.2. Just as in Eq. 5, the sum

$$s = \sum_{i=1}^{N} x_i \tag{9}$$

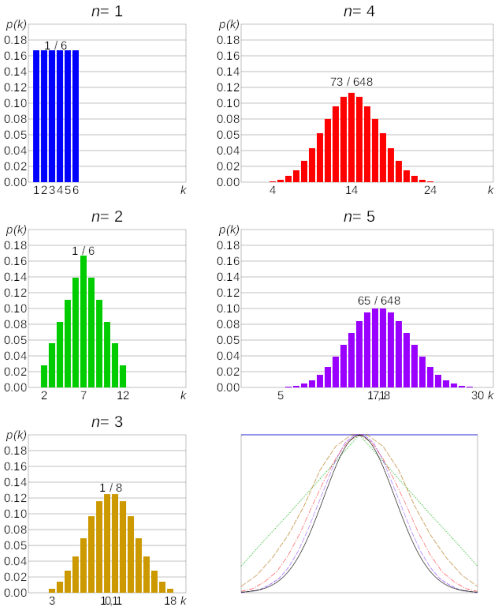
is itself a random variable.

The **central limit theorem** states that for large $N \gg 1$ the probability distribution for s is

$$p(s) \approx \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{(s - N\mu)^2}{2N\sigma^2}\right], \tag{10}$$

with the approximation becoming exact in the $N \to \infty$ limit.

In addition to asserting that the collective behaviour of many independent and identically distributed random variables x_i is governed by a **normal** (or **gaussian**) **distribution**, the central limit theorem further specifies the precise form of this distribution in terms of the mean and variance of *each individual* x_i.



commons.wikimedia.org/wiki/File:Dice_sum_central_limit_theorem.svg

In practice, N often doesn't need to be very large in order for the central limit theorem to provide a reasonable approximation. This is illustrated by the figure to the left, from Wikimedia Commons, which shows the probabilities for the sum of $n \leq 5$ rolls of a six-sided die rapidly approaching a gaussian distribution.

1.5 Diffusion and the central limit theorem

1.5.1 Random walk on a line

As a more powerful application and illustration of the central limit theorem, let's consider the behaviour of a randomly moving object. Such **random walks** appear frequently in mathematical modelling of stochastic phenomena (including Brownian motion), and can be applied to movement through either physical space or more abstract vector spaces. They are examples of Markov processes, in which the state of the system (in this case the position of the 'walker') at any time probabilistically depends only on the system's prior state at the previous point in time — there is no 'memory' of any earlier states. The resulting sequence of system states is known as a *Markov chain*, since each state is produced from the one before, like links in a chain.

Let's consider the simple example of a random walker that moves only in a single spatial dimension — to the left or to the right on a line — and can only take 'steps' of a fixed length, which we can set to $\ell = 1$ without loss of generality. At each point in time, the walker takes either a step to the right (R) with probability p or a step to the left (L) with probability $q = 1 - p$. We will further assume that each step takes a constant amount of time δt, so a walk of N steps will last for total time

$$t = N\delta t. \tag{11}$$

As an example, for $N = 6$ a representative walk can be written as $LRLRRR$, which leaves the walker $x = 2$ steps to the right of its starting point ($x = 0$). The opposite walk $RLRLLL$ would leave the walker at $x = -2$, with negative numbers indicating positions to the left of the starting point. How many possible walks are there for $N = 6$, and what is the probability (in terms of p and q) for the walks $LRLRRR$ and $RLRLLL$ to occur? How many possible walks are there for general N, and what is the probability for any particular walk involving r steps to the right to occur?

We are interested in the walker's final position x at time t after it has taken N steps. Just as for the sum of n rolls of a die considered in Section 1.4, there are a range of possible final positions x, each of which has some probability $P(x)$ of being realized. The key pieces of information we want to determine are the expectation value $\langle x \rangle$ and the standard deviation $\Delta x \equiv \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$ that indicates the scale of fluctuations we can expect around $\langle x \rangle$ as the N-step walk is repeated many times from the same starting point. (We reserve the variables μ and σ^2 for the mean and variance *of the single-step process*, which will play an important role when we apply the central limit theorem in Section 1.5.3.)

Suppose the N total steps involve r steps to the right. What is the final position x of the walker in terms of N and r? Check your general answer for the cases $N = 6$ and $r = 4, 2$ considered above.

This relation makes it equivalent to consider either the probability P_r of taking r steps to the right, or the probability $P(x)$ of ending up at final position x. This equivalence **will not hold** for more general random walks in which the step length is no longer fixed and ℓ_i can vary from one step to the next.

Because the order in which steps are taken does not affect the final position x, to determine the probability $P(x)$ we have to count all possible ways of walking to x. For $N = 6$, what are all the possible walks that produce $x = 4$, and what is the corresponding probability $P(4)$?

Your answer should have a factor of 6 that corresponds to the binomial coefficient $\binom{N}{r} = \binom{6}{5} = 6$. In terms of this binomial coefficient, what is the general probability P_r that an N-step walk will include r steps to the right in any order?

Given this probability P_r, we can apply Eqs. 2–3 to find the expectation value $\langle x \rangle$ and the standard deviation Δx. As a first step, what are $\langle x \rangle$ and $\langle x^2 \rangle$ in terms of the expectation values $\langle r^n \rangle = \displaystyle\sum_{r=0}^{N} r^n \, P_r$?

$\langle x \rangle =$

$\langle x^2 \rangle =$

Now we need to calculate the necessary $\langle r^n \rangle$. A powerful way to do this is to define the **generating function**

$$G(\theta) = \sum_{r=0}^{N} e^{r\theta} \, P_r. \tag{12}$$

This approach introduces a parameter θ that we subsequently remove by setting $\theta = 0$. For example, $G(0) = \sum_{r=0}^{N} P_r = 1$. What do you obtain upon taking derivatives of the generating function and then setting $\theta = 0$?

$\left.\dfrac{d}{d\theta} G(\theta)\right|_{\theta=0} =$

$\left.\dfrac{d^n}{d\theta^n} G(\theta)\right|_{\theta=0} =$

For the current case of a fixed-step-length random walk in one dimension, the probabilities P_r produce a simple closed-form expression for the generating functional:

$$G(\theta) = \sum_{r=0}^{N} e^{r\theta} \, P_r = \sum_{r=0}^{N} e^{r\theta} \binom{N}{r} p^r q^{N-r} = \left(e^{\theta} p + q\right)^N, \tag{13}$$

making use of the binomial formula $(a+b)^N = \sum_{i=0}^{N} \binom{N}{i} a^i b^{N-i}$.

It's straightforward to take the necessary derivatives of Eq. 13, which simplify pleasantly since $\left(e^\theta p + q\right)\big|_{\theta=0} = p + q = 1$:

$$\frac{d}{d\theta}\left(e^\theta p + q\right)^N\bigg|_{\theta=0} =$$

$$\frac{d^2}{d\theta^2}\left(e^\theta p + q\right)^N\bigg|_{\theta=0} =$$

Insert the resulting $\langle r\rangle$ and $\langle r^2\rangle$ into the relations derived above:

$$\langle x\rangle =$$

$$(\Delta x)^2 = \langle x^2\rangle - \langle x\rangle^2 =$$

In the end, you should obtain

$$\langle x\rangle = N(2p-1) \qquad\qquad \Delta x = 2\sqrt{Npq}. \tag{14}$$

We can check that this $\langle x\rangle$ produces the expected results in the special cases $p = 0$, $1/2$ and 1, while the standard deviation Δx also behaves appropriately by vanishing for both $p = 0$ and 1.

1.5.2 Law of diffusion

It's possible to interpret the results in Eq. 14 in a more intuitive way by expressing them in terms of the total time t taken by the random walk (Eq. 11). Inserting $N = t/\delta t$ into Eq. 14,

$$\langle x\rangle = \frac{t}{\delta t}(2p-1) = \frac{2p-1}{\delta t}t \equiv v_{\text{dr}}t,$$

we see that the *expected* final position of the walker depends linearly on time, with **drift velocity**

$$v_{\text{dr}} = \frac{2p-1}{\delta t} = \frac{N(2p-1)}{t} = \frac{\langle x\rangle}{t}. \tag{15}$$

The sign of v_{dr} indicates whether the walker is drifting to the right ($p > \frac{1}{2}$) or to the left ($p < \frac{1}{2}$). The typical scale of fluctuations (or the 'uncertainty') around the expected final position $\langle x \rangle$ is given by the standard deviation Δx, and also depends on time:

$$\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = 2\sqrt{Npq} = 2\sqrt{\frac{pq}{\delta t}}\sqrt{t},$$

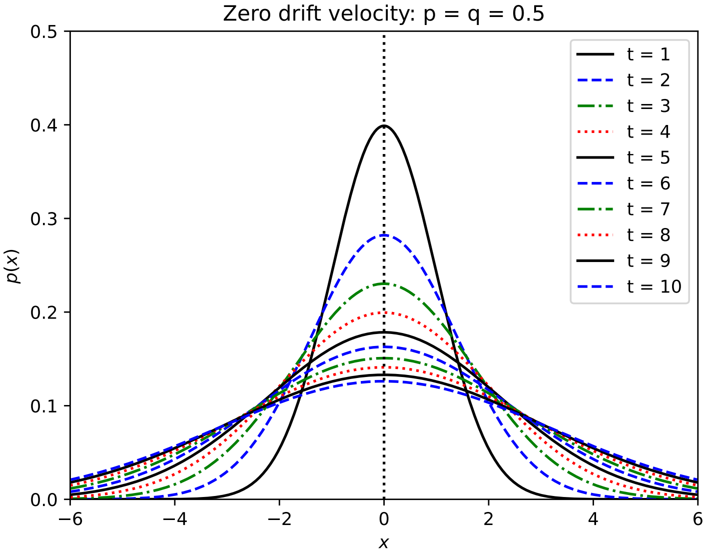
where the constant factor is non-negative. This $\sqrt{t}$ dependence is a very general result.

The **law of diffusion** states that

$$\Delta x = D\sqrt{t}, \tag{16}$$

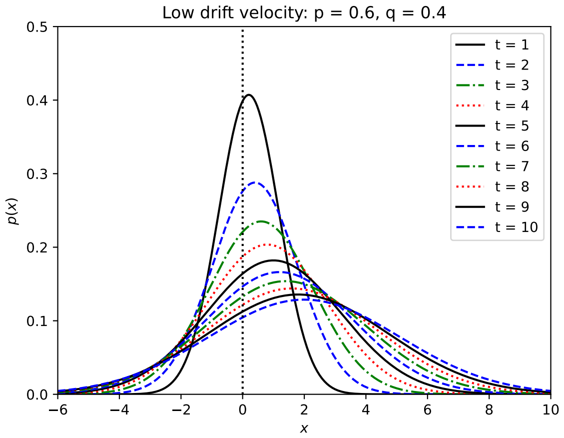
where $D > 0$ is the **diffusion constant**. The standard deviation Δx is sometimes also called the *diffusion length*.

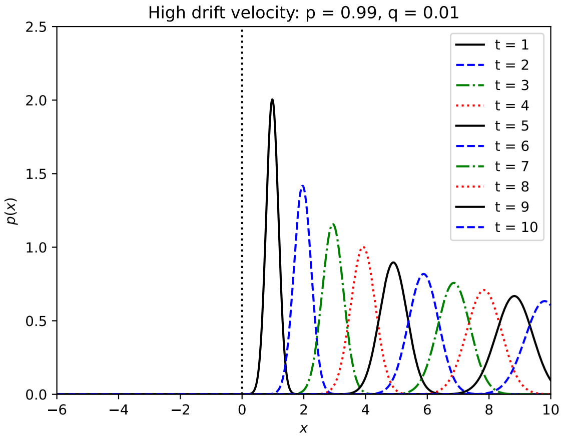
The result $D = 2\sqrt{\frac{pq}{\delta t}}$ that we computed above is specific to the simple example of a fixed-step-length random walk in one dimension and **will not hold** for more general random walks. The behaviour it describes is illustrated by the figure below, which plots the t-dependent probability distribution $p(x)$ that we'll soon derive using the central limit theorem (Eq. 17). What we can see already, even before completing that derivation, is that the probability distribution steadily spreads out — or *diffuses* — as time passes:



Here we are considering the special case $p = q = \frac{1}{2}$, for which the drift velocity $v_{\mathrm{dr}} = 0$ and the expectation value is always $\langle x \rangle = 0$ for any walk time t. However, as time goes on, there is a steady decrease in the probability that the walker will end up around its starting point $x \approx 0$. (As described in Section 1.3, we can determine this probability by integrating the distribution $p(x)$ over an appropriate interval.) Instead, the interval $-D\sqrt{t} \leq x \leq D\sqrt{t}$ within which we can expect to find the walker (with a constant 'one-sigma' or $\sim$68% probability) steadily grows with characteristic dependence on the square root of the time the diffusive process lasts.

Except in the trivial cases $p = 0$ or $q = 0$, diffusion also occurs when the drift velocity is non-zero. This is shown in the two figures below, considering a low but non-zero drift velocity $v_{\text{dr}} = 0.2$ on the left, and a high $v_{\text{dr}} = 0.98$ on the right. All three figures were produced by this Python script.





In the figure on the left, the large diffusion constant $D \approx 0.980$ produces a plot that looks similar to the $v_{\text{dr}} = 0$ case on the previous page, but now the central peaks (and expectation values $\langle x \rangle$) for each t drift steadily to the right. The lower $D \approx 0.199$ for the figure on the right leads to less overlap among the distributions, but they still diffuse to exhibit lower and broader peaks as time passes.

When $p \neq \frac{1}{2}$ so that $\langle x \rangle \neq 0$, it is interesting to compare the drift in the expectation value against the growth in fluctuations around $\langle x \rangle$ due to diffusion. We can do this by considering the following *relative* uncertainty:

$$\frac{\Delta x}{\langle x \rangle} =$$

You should find that at large times this ratio vanishes proportionally to $1/\sqrt{t} \propto 1/\sqrt{N}$. Although the absolute uncertainty $\Delta x = D\sqrt{t}$ grows by diffusion, for $v_{\text{dr}} \neq 0$ the linear drift in the expectation value becomes increasingly dominant as time goes on.

1.5.3 Applying the central limit theorem

Based on our work in Section 1.4, we can see how to apply the central limit theorem to analyse this fixed-step-length random walk in one dimension, for large numbers of steps N or equivalently large times $t = N\delta t$. Each step in the random walk is an independent and identically distributed random variable x_i. The corresponding probability space involves only two possible outcomes: a step of length $\ell = 1$ to the right or to the left with probability p or q, respectively. From this we can easily compute the mean and variance *of the single-step process*:

$\mu = \langle x_i \rangle =$

$\langle x_i^2 \rangle =$

$\sigma^2 = \langle x_i^2 \rangle - \langle x_i \rangle^2 =$

The final position x of the walker after N steps is exactly the sum over these x_i given in Eq. 9. Its probability distribution $p(x)$ from the central limit theorem is therefore obtained directly from these single-step μ and σ^2, which we can also express in terms of the drift velocity and diffusion constant:

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi(4Npq)}} \exp\left[-\frac{(x - N(2p-1))^2}{8Npq}\right] \\ &= \frac{1}{\sqrt{2\pi D^2 t}} \exp\left[-\frac{(x - v_{\text{dr}}t)^2}{2D^2 t}\right]. \end{aligned} \tag{17}$$

This expression was used to produce the three figures above. We could have jumped straight to the final line by considering Eq. 10 and noting

$$v_{\text{dr}}t = N(2p-1) = N\mu \qquad\qquad D^2 t = 4pq\frac{t}{\delta t} = N\sigma^2. \tag{18}$$

While this dependence on p and q is specific to the particular fixed-step-length random walk we're currently considering, the generic results $v_{\text{dr}}t = N\mu$ and $D^2 t = N\sigma^2$ in Eq. 18 **do hold** for any random walk with finite μ and σ^2 (so that the central limit theorem can be applied). This is remarkable, because it means that the diffusive process as a whole is determined entirely by the single-step mean and variance. So long as μ and σ^2 are finite, we end up with Eq. 17 as the large-t probability distribution for any markovian random walk in a single variable x.

This result is related to the generality of the law of diffusion (Eq. 16), which we can recognize in the structure of Eq. 17. Since $t > 0$, the exponential in the gaussian distribution $p(x)$ peaks at the drifting expectation value $x = v_{\text{dr}}t = \langle x \rangle$. The factor

$(x - v_{\mathrm{dr}}t)^2$ simply quantifies the distance from this peak. As t increases, so does the factor $2D^2t$ dividing this $(x - v_{\mathrm{dr}}t)^2$, meaning that a larger distance from the peak is needed for the overall argument of the exponential to reach a given value — in other words, the peak becomes broader. This in turn requires a lower peak, reflected in the $\frac{1}{\sqrt{t}}$ in the overall coefficient, which is set by requiring $\int p(x)\, dx = 1$. In other words, the law of diffusion holds whenever the central limit theorem is applicable. This requires that the mean and variance of the single-step process are finite, and in the computer project we will numerically investigate the *anomalous diffusion* that occurs when this condition is not satisfied.

Unit 2: Micro-canonical ensemble

2.1 Statistical ensembles and thermodynamic equilibrium

We begin this unit by establishing the concept of *statistical ensembles*, which was formalized by J. Willard Gibbs in 1902. (Gibbs also introduced the term 'statistical mechanics', in 1884.) Building on the probability foundations laid above, we will be interested in 'experiments' that simply allow a collection of degrees of freedom to evolve in time, subject to certain constraints. At a given time t_1, the arrangement of these degrees of freedom defines the state ω_1 of the system.

As a concrete example, consider a system of *spins* — arrows that can point either 'up' or 'down' — arranged in a line. Such spin systems will appear several times in the remainder of this module, since in addition to obeying simple mathematics analogous to flipping coins, spins also serve as good models of physical systems such as magnetic molecules. What would be a representative state (or *configuration*) for a system of $N = 8$ spins? How many distinct states are there for this system?

At a different time t_2, the system's state ω_2 is generally different from ω_1. However, there are some measurements we can perform that always produce the same outcome even as the system's state changes over time. These measurements define *conserved quantities*, such as the number of spins considered in the example above.

Another important conserved quantity is the *internal* energy E of an isolated (or 'closed') system,

$$E(\omega_1) = E(\omega_2).$$

The conservation of energy is presumably a familiar concept, and you may also know that it can be rigorously proven through Emmy Noether's first theorem.[4] Because conservation of energy was empirically observed long before Noether's theorem was proven, it also has a more grandiose name: the **first law of thermodynamics**. Another way of stating the first law is that any change in the internal energy of one particular system Ω must be matched by an equal and opposite change in the energy of some other system with which Ω is in contact. This will be important when we consider thermodynamic cycles later in the term.

For now, let's return to the example above, and endow the spin system with an internal energy by placing it in a 'magnetic field' of strength H. That is, if a spin is

---

[4] There are complications when considering the dynamical space-time of general relativity, but that's beyond the scope of this module.

parallel to the field, it contributes energy $-H$ to the total energy E of the system. If a spin is anti-parallel to the field, it instead contributes energy H. For later convenience, we define a positive magnetic field $H > 0$ to point upward, and also define n_+ to be the number of spins pointing upward — parallel to the field and therefore contributing *negative* energy. Similarly, the remaining $n_- = N - n_+$ downward-pointing spins are anti-parallel to the field and contribute positive energy. What is the total energy E of the system in terms of n_+ and n_-? What is E for the representative 8-spin state you wrote down above? What fraction of the states of the spin system have this energy?

If instead we consider $N \sim 10^{23}$ hydrogen (H_2) molecules in a container, we can write a simple expression for the internal energy E by treating each molecule as a *point-like particle*, with no size or structure. In this case each molecule contributes only its kinetic energy, and

$$E = \frac{m}{2}\sum_{n=1}^{N} \vec{v}_n^{\,2} = \frac{1}{2m}\sum_{n=1}^{N} \vec{p}_n^{\,2},$$

where $\vec{v}_n$ is the velocity of the nth molecule, $\vec{p}_n = m\vec{v}_n$ is its momentum, and all molecules have the same mass m.

As forecast at the start of the module, we treat the time evolution of any given system as a stochastic process in which the system probabilistically adopts a sequence of states $\omega_i \in \Omega$:

$$\omega_1 \longrightarrow \omega_2 \longrightarrow \omega_3 \longrightarrow \omega_4 \longrightarrow \cdots$$

This approach is a matter of practicality rather than one of principle. In principle, Newton's laws would allow us to predict the exact time evolution of (say) 10^{23} hydrogen molecules, but only by specifying 10^{23} initial conditions and solving 10^{23} differential equations. Since we cannot hope to record so much information or carry out so many computations, we instead apply probability theory in order to analyse these systems.

This leads us to the following core definition: A **statistical ensemble** is the set of all states $\Omega = \{\omega_1, \omega_2, \cdots\}$ that a system can possibly adopt through its time evolution. Each state ω_i has some probability p_i of being adopted by the system, so we can recognize a statistical ensemble as a probability space.

Because these states ω_i depend on the 'microscopic' degrees of freedom that compose the overall system, we will refer to them as **micro-states** from now on. From the definition of probability in Section 1.1, we have the requirement $\sum_i p_i = 1$, which simply means that the system must be in *some* micro-state at any point in time. The fact that time evolution cannot change any conserved quantities, as discussed above, means that such conserved quantities characterize statistical ensembles. Throughout the next seven units we will consider different statistical ensembles with different sets of conserved quantities.

First we define the **micro-canonical ensemble** to be a statistical ensemble characterized by conserved internal energy E and conserved number of degrees of freedom N — which we will call **particle number** for short.

According to the discussion above, this means that a system governed by the micro-canonical ensemble is *isolated* in the sense that it cannot exchange energy or particles with any other system.

Now that the micro-canonical ensemble is defined, we can connect it to our intuition from everyday physical systems. Let's consider a collection of particles moving around and bouncing (or '*scattering*') off each other in a sealed container. To a first approximation, this should describe the behaviour of air in a room, which our lived experience indicates is spread quite uniformly throughout the room in a way that is stable as time passes. We do not expect all the air in a room to be concentrated in any one corner, nor do we expect strong collective gusts of wind without some clear external influence.

These qualitative expectations illustrate the idea of **thermodynamic equilibrium**, an axiomatic concept in statistical mechanics.[5] We can mathematically define thermodynamic equilibrium through the probabilities p_i that appear in the micro-canonical ensemble.

A micro-canonical system Ω with M micro-states ω_i is in thermodynamic equilibrium if and only if all probabilities p_i are equal. If M is finite, the requirement $\sum_i p_i = 1$ implies

$$p_i = \frac{1}{M}. \tag{19}$$

The full meaning and significance of this definition are not immediately obvious, and we will continue exploring them through consideration of derived quantities such as entropy and temperature. First, it's important to emphasize that this equilibrium is *dynamic*: There is not a single 'equilibrium micro-state' that the system sits in. Instead, the equilibrium system continues probabilistically adopting all possible micro-states as it evolves in time.

---

[5] Our expectation that physical systems generically evolve towards thermodynamic equilibrium as time passes is more formally expressed as the ergodic hypothesis.

2.2 Entropy and its properties

2.2.1 Definition of entropy

We can gain further insight into thermodynamic equilibrium by considering a famous derived quantity.

The **entropy** of a statistical ensemble Ω with a countable number of micro-states M is defined to be

$$S = -\sum_{i=1}^{M} p_i \log p_i, \tag{20}$$

where p_i is the probability for micro-state ω_i to occur. Unless otherwise specified, "$\log$" indicates the natural logarithm with base e.

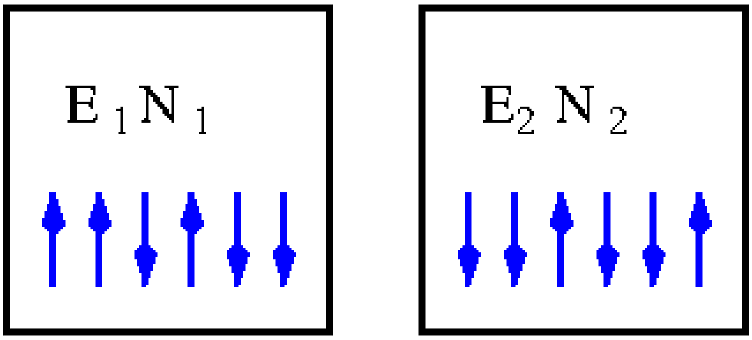
When the system under consideration is in thermodynamic equilibrium, we expect derived quantities such as the entropy to be stable over time, even as different micro-states are probabilistically adopted. This implies that such derived quantities are functions of the conserved quantities that are the same for all micro-states. Therefore, for the micro-canonical ensemble, the equilibrium entropy $S(E, N)$ is a function of the conserved energy and particle number.

By inserting Eq. 19 into Eq. 20 you can quickly compute a simple expression for the entropy of a micro-canonical ensemble in thermodynamic equilibrium:

Your result should depend only on the number of micro-states M, diverging as $M \to \infty$. While the energy E and particle number N are not explicit in this expression, $\{E, N, M\}$ are inter-related and can be expressed in terms of each other given the details of any specific situation under consideration. For example, what is the equilibrium entropy of the system of N spins considered above, if the magnetic field is turned off, $H = 0$? What is the entropy if $E = 0$ with $H > 0$ (which requires $n_+ = n_-$)?

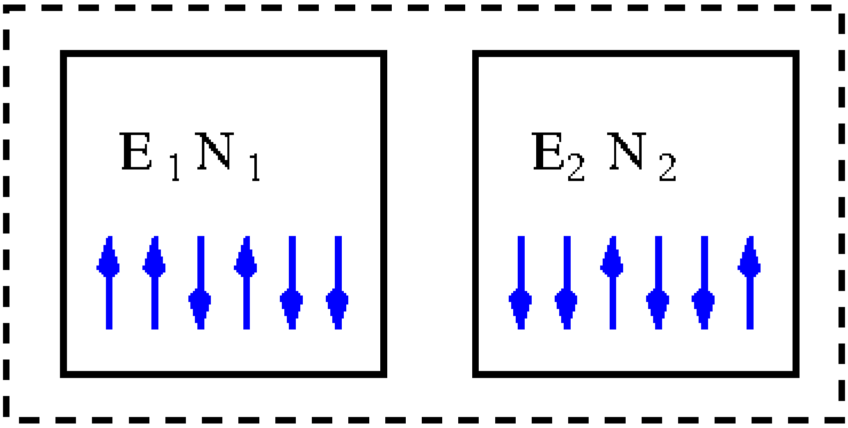
2.2.2 Extensivity

The increase in entropy for an increasing number of micro-states M is a reflection of entropy being an *extensive* quantity. Extensive quantities are formally defined by considering how they behave if two isolated subsystems are *analysed* as a single system — while still remaining isolated from each other, exchanging neither energy nor particles. This is clearest to consider through the specific example shown below of two isolated spin subsystems, Ω_1 & Ω_2, characterized by the energies E_1 & E_2 and particle numbers N_1 & N_2, respectively. To simplify the subsequent analysis, we can assume that both subsystems are placed in magnetic fields with the same H, so that $E_S = -H\left(n_+^{(S)} - n_-^{(S)}\right)$ for $S \in \{1, 2\}$.



We can take system Ω_1 to have M_1 micro-states with probabilities p_i while system Ω_2 has M_2 micro-states with probabilities q_k. As discussed above, each M_S is determined by E_S and N_S. The entropies of the two systems are

$$S_1 = -\sum_{i=1}^{M_1} p_i \log p_i \qquad\qquad S_2 = -\sum_{k=1}^{M_2} q_k \log q_k.$$



Now we keep these two subsystems isolated from each other, but consider them as a combined system Ω_{1+2}, as illustrated above. In order to compute the entropy S_{1+2}, we need to figure out the number of micro-states M_{1+2} the combined system could possibly adopt, and then determine the corresponding probability for each micro-state. Both steps are simplified by the subsystems being isolated from each other, so that they are statistically independent. Specifically, with subsystem Ω_1 in any one of its M_1 micro-states $\omega_i^{(1)}$, subsystem Ω_2 could independently adopt any of its M_2 micro-states, implying $M_{1+2} = M_1 M_2$.

Similarly, statistical independence means that the combined probability of subsystem Ω_1 adopting micro-state $\omega_i^{(1)}$ while subsystem Ω_2 adopts $\omega_k^{(2)}$ is the product of the individual probabilities, $p_i q_k$. We can check this is a well-defined probability, with

$$\sum_{M_{1+2}} p_i q_k = \sum_{i=1}^{M_1} \sum_{k=1}^{M_2} p_i q_k = \left[\sum_{i=1}^{M_1} p_i\right] \cdot \left[\sum_{k=1}^{M_2} q_k\right] = 1 \cdot 1 = 1.$$

Inserting the probability $p_i q_k$ into Eq. 20, and recalling $\log(a \cdot b) = \log a + \log b$, what is the combined entropy S_{1+2} of these two independent subsystems?

$S_{1+2} =$

You should find that the total entropy is the sum of the entropies of the two isolated subsystems, which is also how the energies and particle numbers behave,

$$E_{1+2} = E_1 + E_2 \qquad\qquad N_{1+2} = N_1 + N_2.$$

This behaviour identifies the energy, particle number and entropy as **extensive** quantities, which are defined to be those that add up across independent subsystems. This can be contrasted with **intensive** quantities, which are defined to be independent of the extent of the system, and hence the same (on average) for subsystems as for the combined system. Temperature and density are everyday examples of intensive quantities, though we will see below that the micro-canonical approach introduces some subtleties. It is possible for quantities to be neither extensive nor intensive, for example the number of micro-states $M_{1+2} = M_1 M_2$.

Finally, suppose that each subsystem is independently in thermodynamic equilibrium, with finite M_1 and M_2, implying

$$p_i = \frac{1}{M_1} \qquad\qquad q_k = \frac{1}{M_2}$$
$$S_1 = \log M_1 \qquad\qquad S_2 = \log M_2.$$

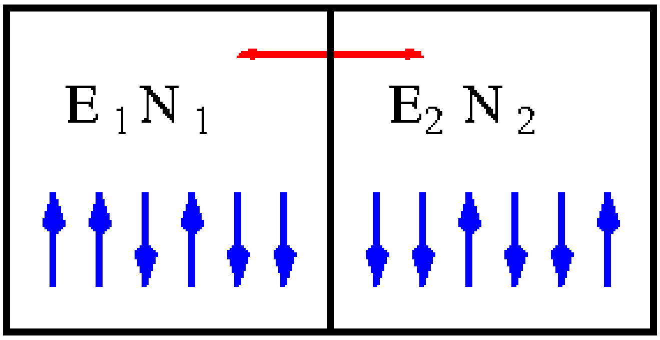
As a consequence we can establish that Ω_{1+2} is also in thermodynamic equilibrium, since the probabilities

$$p_i q_k = \frac{1}{M_1 M_2} = \frac{1}{M_{1+2}}$$

are identical all of its micro-states. In this situation it's even easier to see $S_{1+2} = \log{(M_1 M_2)} = \log M_1 + \log M_2 = S_1 + S_2$.

2.2.3 Second law of thermodynamics

Let's continue considering two subsystems, with one significant change: Suppose the subsystems are now able to exchange energy (but not particles) with each other. We'll say they are in *thermal contact* with each other, rather than being fully isolated. We'll also wait long enough after establishing thermal contact for the combined system Ω to reach equilibrium. This is illustrated below:



The total energy $E = E_1 + E_2$ remains conserved, so the overall system Ω is still governed by the micro-canonical ensemble. However, the individual energies E_1 and E_2 can now change over time, meaning that *each subsystem is no longer micro-canonical*.

The overall system Ω is *not* the same as the combined Ω_{1+2} considered above. We need to reconsider the total number of micro-states M that Ω could adopt, which is much more difficult than before because we can no longer apply statistical independence. Our key remaining tool is the conservation of the total energy E.

Considering a micro-state in which the N_1 spins contribute energy e_1 to the total, we know that the N_2 spins must contribute the remaining $E - e_1$. Our work above implies there are $M_{e_1} = M_{e_1}^{(1)} M_{E-e_1}^{(2)}$ micro-states providing this particular distribution of energies, where $M_{e_1}^{(1)}$ is the number of micro-states of the formerly isolated subsystem Ω_1 with energy e_1, and $M_{E-e_1}^{(2)}$ similarly corresponds to Ω_2 with energy $E - e_1$. We also know that it's possible to have $e_1 = E_1$, since that's the initial energy of Ω_1 before it was brought into thermal contact with Ω_2. When $e_1 = E_1$, we have $M_{E_1} = M_1 M_2$, covering all the micro-states of the combined Ω_{1+2} when the two subsystems were isolated. *In addition*, we also have to count any other micro-states for which $e_1 \neq E_1$:

$$M = \sum_{e_1} M_{e_1}^{(1)} M_{E-e_1}^{(2)} = M_1 M_2 + \sum_{e_1 \neq E_1} M_{e_1}^{(1)} M_{E-e_1}^{(2)} \geq M_1 M_2. \tag{21}$$

Equality holds when $e_1 = E_1$ is the only possibility — this is an extremely special case, in which the two subsystems remain individually micro-canonical, with fixed E_1 and E_2. This is all we can say in full generality, without specifying more details of a particular example, but it allows us to obtain a famous result for the total entropy S of Ω *in thermodynamic equilibrium*:

$$S = \log M \geq \log\left(M_1 M_2\right) = S_{1+2}.$$

This is a form of the **second law of thermodynamics**,

$$S \geq S_{1+2} = S_1 + S_2.$$

In words, whenever initially isolated (sub)systems in thermodynamic equilibrium are brought into thermal contact with each other and allowed to exchange energy, the total entropy of the overall system can never decrease. Indeed, it generically increases except in extremely special cases.

Though we won't go through a more general derivation here, it turns out that the total entropy never decreases (and generically increases) as time passes, under *any* circumstances. This has many far-reaching consequences, the first of which is a more general definition of thermodynamic equilibrium that (unlike Eq. 19) will also apply when we consider statistical ensembles other than the micro-canonical ensemble. For simplicity we assume that any system under consideration has a finite number of micro-states, which means that its entropy is bounded from above. To motivate the definition below, note that the overall system Ω may have undergone an equilibration process to reach its thermodynamic equilibrium after any independently equilibrated subsystems were brought into thermal contact — and in this process the entropy was non-decreasing.

A system is defined to be in **thermodynamic equilibrium** if its entropy is maximal.

We can *derive* Eq. 19 from this definition. All we need to do is maximize the entropy $S = -\sum_i p_i \log p_i$ subject to the three micro-canonical constraints of conserved energy, conserved particle number, and well-defined probabilities $\sum_i p_i = 1$. It turns out that only the final constraint needs to be incorporated into the maximization, through the method of **Lagrange multipliers**. As a reminder, this method involves maximizing the modified entropy

$$\overline{S}(\lambda) = S + \lambda\left(\sum_{i=1}^{M} p_i - 1\right) = -\sum_{i=1}^{M} p_i \log p_i + \lambda\left(\sum_{i=1}^{M} p_i - 1\right),$$

and subsequently imposing $\sum_i p_i = 1$. Here λ is a parameter called the 'multiplier'. In short, this procedure is valid because $\dfrac{\partial \overline{S}}{\partial \lambda} = 0$ once we impose $\sum_i p_i = 1$, so that any extremum of $\overline{S}$ corresponds to an extremum of $S = \overline{S}(\lambda = 0)$.

Recalling $\dfrac{\partial}{\partial x_k}\sum_i f(x_i) = \dfrac{\partial f(x_k)}{\partial x_k}$, what is the probability p_k that maximizes the modified entropy $\overline{S}$?

$$0 = \frac{\partial \overline{S}}{\partial p_k} =$$

You should find that p_k is some constant that depends on λ. We don't care about λ; so long as we know p_k is constant, then we must have $p_k = \frac{1}{M}$ in order to satisfy $\sum_k p_k = 1$. As advertised, we recover Eq. 19 from our new definition of thermodynamic equilibrium based on the second law.

2.3 Temperature

In the micro-canonical ensemble, the conserved internal energy and particle number are fundamental, while the temperature (like the entropy) is a derived quantity. As discussed below Eq. 20, in thermodynamic equilibrium such derived quantities are functions of the conserved $\{E, N\}$. In this section we will state the definition of temperature for the micro-canonical ensemble and apply this to a spin system. In the next section we will check that this definition reproduces our expectations from everyday experiences.

In thermodynamic equilibrium, the **temperature** $T(E, N)$ in the micro-canonical ensemble is defined by

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_N . \tag{22}$$

In words, the (inverse) temperature is set by the dependence of the entropy on the internal energy for a fixed number of degrees of freedom.

Since this definition is not terribly intuitive, we will again gain insight by considering N spins in a line, in a magnetic field of strength H. We saw above that $E = -H(n_+ - n_-)$ for n_+ and $n_- = N - n_+$ spins respectively pointing up and down. With N fixed, each (conserved) value of E defines a *different* micro-canonical system, which we can expect to have a different number of micro-states $M(E)$, different entropy $S(E)$ and different temperature $T(E)$. We will compute the functional forms of each of these three quantities, starting with $M(E)$.

Even though the total energy E remains fixed as time passes, individual spins can ‘flip’ between pointing up or down. Such spin flips simply have to come in pairs so that the overall $n_{\pm}$ both remain the same. As illustration, what are representative spin configurations that produce the minimal energy $E_{\text{min}} \equiv E_0$ and the next-to-minimal E_1? What are E_0 and E_1 in terms of $\{N, H\}$, and how many distinct micro-states are there for each of E_0 and E_1?

Your results should generalize to

$$M(E_{n_-}) = \binom{N}{n_-} = \frac{N!}{n_-!\,(N-n_-)!} = \binom{N}{n_+}. \tag{23}$$

To take the derivative in Eq. 22, we need to express n_- in terms of $\{E, N\}$. It will also be useful to avoid the factorial operation, which is inconvenient to differentiate. For $N \gg 1$, we can accomplish both these goals by treating the spin system as a random walk in the space of its possible energies E and applying the central limit theorem:[6]

- Each spin adds to $x \equiv \frac{E}{-H} = 2n_+ - N$ a ‘step’ of fixed ‘length’ ± 1. Our task therefore coincides with the special case we considered in Section 1.5.
- We don’t impose any preference for positive vs. negative energies, meaning $p = q = \frac{1}{2}$ in the terminology of Section 1.5.
- With $p = q = \frac{1}{2}$, every one of the 2^N possible configurations of N spins is equally probable. Therefore the probability P_{n_+} that our overall ‘walk’ ends up producing a configuration with $n_+ = \frac{1}{2}(x + N)$ is simply the fraction of those 2^N states with this n_+, in which we can recognize Eq. 23:
$$P_{n_+} = \frac{1}{2^N}\binom{N}{n_+} = \frac{M(E_{n_-})}{2^N} \qquad \Longrightarrow \qquad M(E_{n_-}) = 2^N P_{n_+}.$$
- To estimate P_{n_+} for $N \gg 1$, we apply the central limit theorem just as in Section 1.5.3. In particular, we can re-use our computation that $\mu = 2p - 1 = 0$ and $\sigma^2 = 4pq = 1$, to find
$$p(x) \approx \frac{1}{\sqrt{2\pi N}} \exp\left[-\frac{x^2}{2N}\right].$$
This is the probability *distribution* from which we want to extract P_{n_+}.

---

[6] Applying Stirling’s formula, $\log(N!) \approx N \log N - N$, is another possible approach.

- From a tutorial problem we know that $P_{\text{const}}(n_+) = p(2n_+ - N)\Delta n_+$ is a good approximation. With $\Delta n_+ = 1$ and $2n_+ - N = \frac{E}{-H}$, we therefore find

$$M(E) \approx 2^N p(2n_+ - N) = \frac{2^N}{\sqrt{2\pi N}} \exp\left[-\frac{E^2}{2NH^2}\right]. \tag{24}$$

What is the derivative of the log of Eq. 24 with N fixed?

$$\left.\frac{\partial}{\partial E} \log M\right|_N =$$

You should find the temperature

$$T \approx -\frac{NH^2}{E} \qquad\qquad N \gg 1, \tag{25}$$

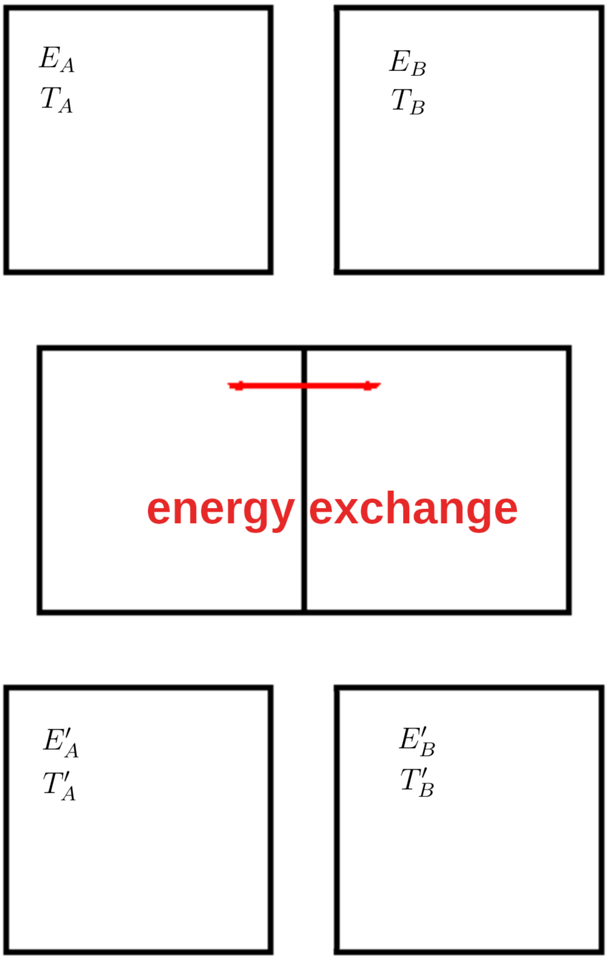
which in several ways does *not* seem to match our expectations from everyday experiences: This T diverges as $E \to 0$ for $n_+ \approx n_-$, and it is negative whenever $n_+ < n_-$ to produce $E > 0$. You can check that this $T < 0$ corresponds to the number of micro-states decreasing for larger internal energies, $\frac{\partial M}{\partial E} < 0$. In **natural** systems, larger energies make more micro-states accessible, producing $\frac{\partial M}{\partial E} > 0$ and a positive temperature. If $H = 0$, we also have $E = 0$ and T is ill-defined.

Restricting our attention to $H > 0$ and $n_+ > n_-$, we also see that the resulting non-negative temperature cannot vanish. It is minimized by the most-negative energy you found above, $T_{\text{min}} = H > 0$ for $E_{\text{min}} = -NH$. The non-zero minimum temperature is specific to spin systems, while some of the other oddities result from the micro-canonical approach more generally. This will motivate turning to the canonical ensemble in Unit 3, but first we can check that some aspects of the micro-canonical temperature defined in Eq. 22 do match our everyday expectations, at least in the 'natural' positive-temperature regime.

2.4 Heat exchange

From Eq. 25 for the temperature of a micro-canonical spin system, we can see that 'natural' positive temperatures correspond to negative energies, and therefore increase as the energy increases by becoming less negative (with a smaller magnitude). Such a direct relation between energy and temperature is very generic, and we will study it in more detail when considering thermodynamic cycles in a few weeks. For now, considering unspecified systems that exhibit this natural behaviour, let's ask what would happen if we take two initially isolated micro-canonical systems — Ω_A and Ω_B with temperatures T_A and T_B in thermodynamic equilibrium — and bring them into thermal contact.

In micro-canonical terms, the temperatures T_A and T_B are derived from the corresponding energies E_A and E_B, while thermal contact allows the two systems to exchange energy (but not particles) as non-isolated subsystems of a combined micro-canonical system Ω_C. Once the two subsystems have been in thermal contact long enough for the combined system to have reached thermodynamic equilibrium, it will have temperature T_C. We can then re-isolate the two subsystems, which will remain in thermodynamic equilibrium with energies $\{E'_A, E'_B\}$ and temperatures $\{T'_A, T'_B\}$. This three-step procedure is illustrated below.



From everyday experience, we expect that this energy exchange will result in a net flow of energy from the hotter system to the colder system, cooling the former by heating the latter. We will now check that the micro-canonical definition of temperature in Eq. 22 predicts this expected behaviour. With $S \in \{A, B\}$, we can write

$$E'_S = E_S + \Delta E_S$$

and consider for simplicity the case where the change in energy is relatively small,

$$\left|\frac{\Delta E_S}{E_S}\right| \ll 1.$$

Since we can build up large changes in energy through a series of smaller changes, this assumption doesn't lead to any loss of generality. We also know $\Delta E_B = -\Delta E_A$ thanks to conservation of energy.

Equation 22 tells us that we need to consider the entropies as functions of E_S and E'_S in order to connect the temperatures to any flow of energy. Because we don't change the number of particles in each system, we only need to consider the energy dependence of the entropy. We assume $S(E)$ is continuous and infinitely differentiable,[7] which allows us to expand each of the final entropies $S(E'_S)$ in a Taylor series,

$$S(E'_S) = S(E_S + \Delta E_S) \approx S(E_S) + \left.\frac{\partial S}{\partial E}\right|_{E_S} \Delta E_S,$$

neglecting all $\mathcal{O}(\Delta E_S^2)$ terms because we consider relatively small changes in energy. What is the expression above in terms of the initial temperatures T_S?

From the second law of thermodynamics, we know that the total entropy of these systems can never decrease as time passes:

$$S(E_A) + S(E_B) \leq S(E_A + E_B) = S(E'_A) + S(E'_B). \tag{26}$$

The final equality means that re-isolating the two subsystems doesn't change the entropy. This is because E'_A is not fixed and could take any value from zero to $E_A + E_B$ at the moment when the subsystems are re-isolated. Computing the final entropy $S(E'_A) + S(E'_B)$ therefore requires summing over all possible values of E'_A, producing exactly the sum in Eq. 21 for the overall system. We will see something similar when we consider the 'Gibbs paradox' in Unit 4.

What do you find when you insert your linearized Taylor series into Eq. 26?

---

[7] This assumption breaks down at a *phase transition*, where we would need to be more careful. We will learn about phase transitions towards the end of the term.

Applying conservation of energy should produce

$$\left(\frac{1}{T_A} - \frac{1}{T_B}\right) \Delta E_A \geq 0.$$

Recalling from Section 2.2.3 that equality holds only in extremely special cases, we can identify three possibilities consistent with this result. If $T_A > T_B$, then $\left(\frac{1}{T_A} - \frac{1}{T_B}\right)$ is negative and we will generically have $\Delta E_A < 0$, so that energy flows out of the hotter system Ω_A and into the colder one. Our restriction to natural systems means this flow of energy reduces the higher temperature, and increases the lower temperature, bringing the temperatures of the two subsystems closer to each other. Similarly, if $T_A < T_B$, we will generically have $\Delta E_A > 0$, meaning that energy still flows from the hotter system Ω_B into the colder one, again reducing the difference in their temperatures. We can finally conclude that $T_A = T_B$ is the very special case where there is no energy flow, $\Delta E_S = 0$, keeping the temperatures the same. All of this is exactly what we would expect based on our everyday experience of temperature as an intensive quantity.