Introduction

These notes give an introduction to statistical mechanics, and explore how the tools of statistical mechanics can be used to describe complex systems.

To understand the words in the title of the lecture, a *complex systems* typically means a collection of many interconnected elements, which display collective (or emergent) phenomena not trivially extrapolated from the behaviour of the individual constituents.

To have an illustration about *statistical mechanics*, let's consider a physical system made of a number of interacting particles. When it is just a single particle in given potential, it is an easy problem: one can write down the solution (even if one could not calculate everything in closed form). Having 2 particles is equally easy, as this so-called "two-body problem" can be reduced to two modified one-body problems (one for the centre of mass, other for the relative position). However, a dramatic change occurs at 3 particles. The study of the three-body problem started with Newton, Lagrange, Laplace and many others, but the general form of the solution is still not known. Even relatively recently, in 1993 a new type of periodic solution has been found, where 3 equal mass particles interacting gravitationally chase each other in a figure-8 shaped orbit. This and other systems where the degrees of freedom is low belongs to the subject of dynamical systems, and is discussed in detail in Chapter 2 of this volume. When the number of interacting particles increases to very large numbers, like 10^{23}, which is typical for the number of atoms in a macroscopic object, surprisingly it gets simpler again, as long as we are interested only at aggregate quantities. This is the subject of statistical mechanics.

Statistical mechanics is also the microscopic foundation of thermodynamics. It developed a number of powerful tools, which can be used outside of the convential physics domain, like biology, finance, traffic, and more. It can also be considered as *the science of ignorance*: how to handle a system where we do not know (even prefer not to know) everything. A particular example is renormalisation theory, developed in the second half of the 20th century, which gives a systematic framework to dispense with successively the non-interesting degrees of freedom.

Our approach is that of a physicist: (i) based on models, (ii) fundamentally quantitative, (iii) but not rigorous. Probably the most important concept in science is that it is possible to construct an abstraction ("model") of the world around us which is admittedly much simpler than the real thing, but nevertheless captures important characteristics which enable to make predictions which are not "explicitly put in" into the model. The models are fundamentally quantitatve, and we use mathematical language to describe them. We stop here however, and leave rigorous treatment to mathematicians — this approach enables one to progress much quicker, on the expense of losing that what we do is absolutely unshakable.

These notes start with elements of information theory, which in turn are used to derive the foundations of statistical mechanics based on the maximum entropy principle. Unlike the typical treatment in Physics textbooks, this approach has the advantage that the abstract formalism developed can be used in a more straightforward way for systems outside the typical range of thermal applications. We will follow by considering the effects of fluctuations to provide a link to thermodynamics. One of the most characteristic collective phenomena of complex systems is phase transition, which we approach from the direction of statistical physics. The second part of the notes deal with dynamics in some form: we will consider granular materials, transport and traffic models, and collective biological motion like flocking.

Introduction to information theory

Random variables

While probability theory can (and should) be founded rigorously, in these notes we take a relaxed approach and attempt define everything without mentioning the probability space. We call a *random variable* an object which can take values when observed, say random variable X can take any values from $x_1, x_2, \ldots, x_n$. When observed many times, x_1 is taken n_1 times, x_2 is taken n_2 times, etc. The *probabilities* of the outcomes can be defined as relative frequencies in the limit of a large number of observations, for example:

$$\frac{n_1}{\sum n_i} \to p_1$$

It immediately follows that the probabilities add up to 1:

$$\sum_{i=1}^{n} p_i = 1$$

We then say the the probability of X taking a given value x_i is p_i:

$$P(X = x_i) = p_i$$

A *function of a random variable* is another random variable: if X takes x_i with probability p_i, then $f(X)$ takes the value $f(x_i)$ with probability p_i.

The *expectation* or *average* of a random variable can be considered as observing it many times and taking the average of the observed values. In the statistical mechanics literature the standard notation is angular brackets:

$$\langle X \rangle = \sum_i x_i p_i, \qquad \text{or} \qquad \langle f(X) \rangle = \sum_i f(x_i) p_i$$

The above concepts can be easily extended to random variables which can take infinitely many discreet values, and even to ones which take values from a continuum. In the latter case eg. if X can take real values, the probability that X takes any value in $[x, x + dx]$ is $p(x)dx$, where dx is small, and $p(x)$ is called the probability density function. The sums above are replaced with integrals, eg. $\int p(x)dx = 1$. While this naive approach to continuous random variables is sufficient for these notes, in general, especially when dealing whith continuous random variables, one needs a rigorous foundation of probability theory.

The above *frequentist* approach to probabilities is not the only one. In a sentence like "Tomorrow we will have a 10% chance for rain", probabilities are interpreted as a degree of belief or confidence.

The information entropy

Suppose we want to describe the outcome of a sequence of coin tosses (heads or tails): HTHHTHTTTH... This seqence looks very different from that of a lottery play: LLLLL...LLLL...LLWLLLL... The second sequence is much more boring, one can describe it as eg. "trial no. 857,923 was a win, all others were lose". In the first sequence, however, we cannot get away much better than quoting the whole sequence verbatim.

To quantify this difference, we introduce the *information entropy H* of a random variable. Its intuitive meaning is the amount of uncertainty in an observation of a random variable, or in other words the amount of information we gain when observing a random variable. One can think about is as the "amount of answers" needed on average to learn the outcome of an observation as a response to an optimally crafted question-tree. It is a function of the probabilities only: $H(p_1, p_2, \dots)$.

We require certain regularity properties:

(i) continuity: $H(p_1, p_2, \dots)$ is a continuous function of its arguments.

(ii) "sense of direction": of the random variables that take all outcomes with equal probability, the ones with more outcomes carry more information: the function

$$h(n) \overset{\text{def}}{=} H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \tag{1}$$

is monotonically increasing with n.

(iii) "consistency": if H is calculated in two different ways, they should agree. Eg. to calculate the information entropy of a 3-state random variable, we can group the last two states and first obtain the information entropy for the obtained 2-state random variable, and then with probability $p_2 + p_3$ need to resolve the grouped states:

$$H_3(p_1, p_2, p_3) = H_2(p_1, q) + q H_2\left(\frac{p_2}{q}, \frac{p_3}{q}\right)$$

where $q = p_2 + p_3$.

It can be shown that this three requirements restrict the functional form of $H(\cdot)$, see [1]:

$$H(q_1, q_2, \ldots, q_r) = -K \sum_{i=1}^{r} q_i \ln(q_i) \tag{2}$$

This is the information entropy of a random variable with probabilities $q_1, q_2, \ldots, q_r$. The constant K sets the units, which can be fused into the logarithm as setting its base. In many of the following formulae we will use the notation

$$H(q_1, q_2, \ldots, q_r) = -\sum_{i=1}^{r} q_i \log(q_i) \tag{3}$$

without explicitly specifying the base of the logarithm. Then $K = 1/\ln(2)$ in (2), or equivalently using $\log_2$ in (3) the information entropy is measured in *bit*s. When setting $K = 1$ or using $\log_e$ the units are *nat*s, and finally the decimal case is $K = 1/\ln(10)$ or using $\log_{10}$, when the units are called *ban*s.

Multiple random variables

When X and Y are random variables, we can look at the probabilities of (X, Y) pairs. These are called *joint probabilities*:

$$p_{ij} = P(X = x_i, Y = y_j)$$

The probability of one of the random variables (the *marginal probabilities*) are obtained by summing up the joint probabilities on all states of the other random variable:

$$p_i^{(X)} = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

and similarly $p_j^{(Y)} = \sum_i p_{ij}$.

Two random variables are called *independent*, if the joint probabilities factorise into marginals for all (i, j) pairs:

$$\text{if} \quad P(X = x_i, Y = y_j) = P(X = x_i)P(Y = x_j) \qquad \text{for all } i, j$$

or equivalently

$$\text{if} \quad p_{ij} = p_i^{(X)} p_j^{(Y)} \qquad \text{for all } i, j$$

The *conditional probabilities* tell the probability of one random variable when we know the value of another:

$$p_{i|j} \stackrel{\text{def}}{=} P(X = x_i \mid Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_j^{(Y)}}$$

The *joint information entropy* is the uncertainty of the (X, Y) pair:

$$H(X, Y) = -\sum_{i,j} p_{ij} \log p_{ij}$$

The *conditional information entropy* gives the uncertainty of X when Y is known:

$$H(X \mid Y) \stackrel{\text{def}}{=} \langle H(X \mid Y = y_j) \rangle_Y = \ldots = H(X, Y) - H(Y)$$

Finally the *mutual information* is defined as

$$I(X; Y) = \sum i, j p_{ij} \log \frac{p_{ij}}{p_i^{(X)} p_j^{(Y)}} = \ldots = H(X) - H(X \mid Y)$$

so its meaning is the reduction in uncertainty of X due to the knowledge of Y, or in other words how much Y tells about X.