Part 2: Basics of Statistical Physics

Probabilities
- Simple probabilities
- Averages
- Composite probabilities
- Independent events
- Simple and complex
- Many dice
- Probability distributions

Statistical Mechanics
- Hamiltonian description
- Averages from derivatives
- One and many
- Structural Invariance
- Intensive and Extensive

Gaussian
- Statistical Variables
- Integrals and Probabilities
- Statistical Distributions
- Averages
- Gaussian random variable
- Approximate Gaussian integrals

Calculation of Averages and Fluctuations
- The Result
- Going Slowly
- Sums and averages in classical mechanics
- More sums and averages
- Homework
Simple probabilities (reprise)

mutually exclusive events described by $\alpha=1,2,3,\ldots$

number of times $\alpha$ turns up=$N_\alpha$; total number of events $N$

$$N = \sum_\alpha N_\alpha$$

probability of getting a side with number $\alpha$ is $\rho_\alpha$

$$\rho_\alpha = N_\alpha/N$$  \hspace{1cm} \text{ii.1}$$

total probability =1 -->

$$\sum_\alpha \rho_\alpha = 1$$  \hspace{1cm} \text{ii.2}$$

relative probability: relative chance that $\alpha$ will turn up =$r_\alpha$, e.g. fair dice have $r_\alpha =$constant

from $r$ to $\rho$

$$z = \sum_\alpha r_\alpha$$

normalize (=fix up size) :

$$\rho_\alpha = r_\alpha/z$$

cubic dice 6 sides: fair dice --> all probabilities are equal -->

$$r_\alpha=1 \rightarrow z=6 \rightarrow \rho_\alpha=1/6$$

for all values of $\alpha$
Do we understand what this formula means?? How would we describe a loaded die? An average from a loaded die? If I told you that $\alpha=2$ was twice as likely as all the other values, and these others were all equally likely, what would be the relative probability? What would we have for the average throw on the die?
Composite Probabilities

$\alpha$ and $\beta$ are two different kinds of events.
$\alpha$ might describe the temperature on January 1, $\rho_\alpha$ computed as $N_\alpha / N$.
$\beta$ might describe the precipitation on December 31, with probabilities $\rho'_\beta$.

Both kinds of events are complete
\[ \sum_\alpha \rho_\alpha = 1 \quad \sum_\beta \rho'_\beta = 1 \]

The prime indicates that the two probabilities are quite different from one another.

Let $\rho_{\alpha,\beta}$ be the probability that both will happen. The technical term for this is a joint probability. The joint probability satisfies
\[ \sum_{\alpha,\beta} \rho_{\alpha,\beta} = 1 \]

$\rho(\alpha|\beta)$ is the probability that event $\alpha$ occurs if we know that event $\beta$ has or will occur. This quantity is called a conditional probability. It obeys $\rho(\alpha|\beta) = \rho_{\alpha,\beta} / \rho'_\beta$.

Something must happen, implies that
\[ \sum_\alpha \rho(\alpha | \beta) = 1 \]
Independent Events

Physically two events are independent if the outcome of one does not affect the outcome of the other. It is a mutual relation, if $\alpha$ is independent of $\beta$ then $\beta$ is independent of $\alpha$.

This can then be stated in terms of conditional probabilities. If $\rho(\alpha|\beta)$ is independent of $\beta$ then we say $\alpha$ and $\beta$ are statistically independent. After a little algebraic manipulation, it follows that the joint probability $\rho_{\alpha,\beta}$ obeys

$$\rho_{\alpha,\beta} = \rho_{\alpha} \rho_{\beta}'$$

equivalently, two events are statistically independent, if the number of times both show up is expressed in terms of the number of times each one individually shows up as

$$N_{\alpha,\beta} = N_{\alpha}N_{\beta}'/N$$

This can be generalized to the statement that a series of $m$ different events are statistically independent if the joint probabilities of the outcomes of all these events is simply the product of all the $m$ individual probabilities.

The word uncorrelated is also used to describe statistically independent quantities.
Simple and Complex

definition: simple outcome: can happen only one way: like 2 coming up when a die is thrown.

definition: complex outcome: can happen several ways: like 7 coming up when two dice are thrown.

One should calculate probability of complex outcome as a sum of probabilities of simple outcomes.

If the simple outcomes are equally likely, probability of complex outcome is the number of different simple outcomes times the probability of a single simple outcome. There is lots of counting in statistical mechanics. The number of ways that something can happen is often denoted by the symbol $W$. Entropy is given by

Entropy $S = k \ln W$, where $k=k_B$ is Boltzmann’s constant.
Many Dice

Given two fair dice
what is the average sum and product of what turns up

Given two dice independent of one another: one fair and the other one unfair. How does one describe the probabilities of the outcome of throwing both dice together?

what is the chance, rolling two fair dice, that we shall roll an eleven, a seven?

Now we roll one hundred dice all at the same time. What is the average of the sum of the dice-values. How can one define a root mean square fluctuation in this value? How big is it?
Probability Distributions

So far we have talked about discrete outcomes. A die may take on one of six possible values. But measured things are often continuous. For example, in one dimension, the probability that a quantum particle will be found between \( x \) and \( x+dx \) is given in terms of the wave function, \( |\psi(x)|^2 dx \). In this context, the squared wave function appears as a probability density. In general, we shall use the notation \( \rho(x) \) for a probability density, saying that \( \rho(x) \, dx \) is the probability for finding a particle between \( x \) and \( x+dx \). The general properties of such probability densities are simple. They are positive. Since the total probability of some \( x \) must be equal to one they satisfy the normalization condition

\[
\int_{-\infty}^{+\infty} \rho(x) \, dx = 1
\]

For example, in classical statistical mechanics, the probability density for finding a particle with \( x \)-component of momentum equal to \( p \) is

\[
\left( \frac{2\pi\beta}{m} \right)^{1/2} \exp\left[ -\beta \frac{p^2}{(2m)} \right]
\]

This is called a Gaussian probability distribution, i.e. one that is based on \( \exp(-x^2) \). Such distributions are very important in theoretical physics.
Statistical Mechanics: Hamiltonian Description (reprise)

The first idea in statistical mechanics is that the relative probability of a configuration labeled by $\alpha$, is given in terms of the Hamiltonian $H$ or equivalently the energy $E(\alpha)$ in that the relative probability of such a configuration is $\exp(-H/T) = \exp[-E(\alpha)/T]$ where $T$ is the temperature measured in energy units. (This is usually written $kT$ where $k$ is the Boltzmann constant.) The general formulation of statistical mechanics is then that

$$\rho_\alpha = e^{-E(\alpha)/T} / Z \quad \text{with} \quad Z = \sum_\alpha e^{-E(\alpha)/T}$$

For example a single atomic spin in a magnetic field has a Hamiltonian which can be written as $H = -B\sigma_\mu$, with $B$ the $z$ component of the magnetic field and $\sigma$ being the $z$ component of the spin. In this model, this component takes on the values $\sigma = \pm 1$. (Most often, one puts subscripts $z$ on the spin and magnetic field.)

The relative probabilities for positive and negative values of the spin are respectively $e^h$ and $e^{-h}$ with $h = B\mu/(kT)$ so that the partition function is

$$Z = e^h + e^{-h} = 2 \cosh h$$

and the probabilities are

$$\rho(\sigma=1) = e^h/Z \quad \text{and} \quad \rho(\sigma=-1) = e^{-h}/Z$$

so that we can calculate

$$\langle \sigma \rangle = \rho(\sigma=1)(1) + \rho(\sigma=-1)(-1) = (e^h + e^{-h})/z = \tanh h$$
Averages from Derivatives (reprise)

\[ z = \sum_\sigma \exp(h\sigma) = 2 \cosh h \]

\[ \frac{d(\ln z)}{dh} = \sum_\sigma \sigma \exp(h\sigma) / z = \langle \sigma \rangle = \tanh h \]

homework: Read Chapters 1 and 2 in textbook.
show that \(-d(\ln Z) / d\beta = E = \langle H \rangle\) and \(d^2(\ln Z) / d\beta^2 = \langle (H - \langle H \rangle)^2 \rangle\) and

\[ \frac{d^2(\ln z)}{(dh)^2} = \sum_\sigma (\sigma - \langle \sigma \rangle)^2 \exp(h\sigma) / z = \langle (\sigma - \langle \sigma \rangle)^2 \rangle \]

\[ = 1 - \langle \sigma \rangle^2 = 1 - (\tanh h)^2 \]

note how the second derivative gives the mean squared fluctuations.

**All** derivatives of the log of the partition function are thermodynamic functions of some kinds. As I shall say below, we expect simple behavior from the log of \(Z\) but not \(Z\) itself. The derivatives described above are respectively called the magnetization, \(M = \langle \sigma \rangle\) and the magnetic susceptibility, \(\chi_s = dM/dH\). The analogous first derivative with respect to \(\beta\) is minus the energy. The next derivative with respect to \(\beta\) is proportional to the specific heat, or heat capacity, another traditional thermodynamic quantity.
One and Many

Imagine a material with many atoms, each with its own spin. The system has a Hamiltonian which is a sum of the Hamiltonian of the different atoms

\[ H = \sum_{\alpha=1}^{N} h \sigma_\alpha \]

and a probability distribution

\[ \rho = \frac{\exp(-\beta H)}{Z} = \frac{(1 / Z) \prod_{\alpha=1}^{N} \exp(h \sigma_\alpha)}{Z} \]

which is a product of pieces which belong to the different atoms. The different pieces are then \textit{statistically independent} of one another. Note that the partition function is

\[ Z = \prod_{\alpha=1}^{N} \sum_{\sigma^\alpha = \pm 1} \exp(h \sigma^\alpha) = (2 \cosh h)^N = z^N \]

so that the entire probability is a product of \( N \) pieces connected with the \( N \) atoms

\[ \rho\{\sigma\} = \prod_{\alpha} [\exp(h \sigma_\alpha) / z] \]

The appearance of a product structure depends only upon having a Hamiltonian which is a sum of terms referring to individual parts of the system

Hamiltonian is sum \( \longleftrightarrow \) stat mech probability is product \( \longleftrightarrow \) statistical independence
Structural invariance

Note how the very same structure which applies to one atom \( \exp(-\beta H)/Z \) carries over equally to many atoms.

This structural invariance is characteristic of the mathematical basis of physical theories. Newton’s gravitational theory seemed natural because the same law which applied to one apple equally applies to an entire planet composed of apples.

This same thing works for electromagnetism. A wave function is the same sort of thing for one electron or many. The structure of space and time has a similar invariance property. Remember that a journey of a thousand miles starts with but a single step. The similarity between a single step and a longer distance is a kind of structural invariance. This invariance of space is called a scale invariance. It is quite important in all theories of space and time.
**Intensive and Extensive**

Imagine a situation in which we have many independent identical subsystems, called *iid* by the mathematicians for independent and identically distributed. Then $Z(\beta) = z(\beta)^N$ as in equation ii.4. In this kind of system, the free energy, the entropy, and the average energy will be precisely $N$ times that for the single subsystem.

Quantities that are a sum or integral over the entire system are called *extensive*. One which define local properties or are ratios of extensive quantities are called *intensive*. The former are almost always linear in $N$, the latter are almost always independent of $N$.

The mean square fluctuation in energy is extensive. Why? The root mean square fluctuation in energy, $\langle (E-\langle E \rangle)^2 \rangle^{1/2}$, is not extensive. Why? In the limiting case in which the number of subsystems (particles) is very large, which is the relative size of the average energy and the root mean square fluctuation in the energy?

I have just told you the reason why classical thermodynamics does not consider fluctuations in things. Again, what is this reason?

How big is $N$, say for the particles in this room?

In terms of real MKS numbers, roughly what is the average kinetic energy of these particles and what is the size of its fluctuation?
Gaussian Statistical Variables

A Gaussian random variable, $X$, is one which has a probability distribution which is the exponential of a quadratic in $X$.

$$\rho(x) = \left[\frac{\beta}{(2\pi)}\right]^{1/2} \exp\left[-\beta(x-<X>)^2/2\right]$$

$1/\beta$ is the variance of this distribution.

The sum of two statistically independent Gaussian variables is also Gaussian. **How does the variance add up?**

A Gaussian variable is an extreme example of a structurally stable quantity.

**Central Limit Theorem:** A sum of a large number of individually quite small random variables need not be small, but that sum is, to a good approximation, a Gaussian variable, given only that the variance of the individual variables is bounded.

A Gaussian distribution has a lot of structurally invariant properties.
Gaussian integrals and Gaussian probability distributions

Gaussian integrals are of the form

\[ I = \int dx \, \exp(-ax^2/2 + bx + c) \]

with \(a, b,\) and \(c\) being real numbers, complex numbers, or matrices. They are very, very useful in all branches of theoretical physics.

We define the probability that the random variable \(X\) will take on the value between \(x\) and \(x+dx\) as \(\rho(X=x)dx\) or more simply as \(\rho(x)dx\).

There is a canonical form for Gaussian probability distributions, namely

\[ \rho(X=x) = \left(\frac{\beta}{2\pi}\right)^{1/2} \exp\left[-\beta(x-<X>)^2/2\right] \]

produced by “completeing the square”. Here \(1/\beta\) is the variance and \(<X>\) is the average of the random variable, \(X\).

For Gaussian probability distributions, there is a very important result:

\[ <\exp(iqX)> = \exp(iq<X>) \exp[-q^2/(2\beta)] \]

prove this

Notice how the \(\beta\) that appears in the numerator of the probability distribution reappears in the denominator of the average.
Gaussian Distributions

According to Ludwig Boltzmann (1844 – 1906) and James Clerk Maxwell (1831-1879) the probability distribution for a particle in a weakly interacting gas as is given by

\[ \rho(p,r) = \frac{1}{z} \exp(-\beta H) \]

\[ H = \left[ p_x^2 + p_y^2 + p_z^2 \right] / 2m + U(r) \]

Here, the potential holds the particles in a box of volume \( \Omega \), so that \( U \) is zero inside a box of this volume and infinite outside of it. As usual, we go after thermodynamic properties by calculating the partition function,

\[ z = \Omega \left[ \int dp \exp(-\beta p^2 / (2m)) \right]^3 = \Omega (2\pi m / \beta)^{3/2} \]

In the usual way, we find that the average energy is \( 3/(2\beta) = (3/2)kT \). The classical result is the average energy contains a term \( 1/2 kT \) for each quadratic degree of freedom. Thus a harmonic oscillator has \( \langle H \rangle = kT \).

Hint for theorists: Calculations of \( Z \) (or of its quantum equivalent, the vacuum energy) are important. Once you can get this quantity, you are prepared to find out most other things about the system.
Gaussian Averages

The usual way

Let one particle be confined to a box of volume $\Omega$. Let $U(r)$ be zero inside the box and $+\infty$ outside. Then, in three dimensions

\[ Z = \int d^3p \ d^3r \ \exp(-\beta[p^2/(2m)+U(r)]) = \Omega (2\pi m/\beta)^{3/2} \]

Let $\epsilon = [p^2/(2m)+U(r)]$

\[ \partial \ln Z / \partial \beta = -(1/Z) \int d^3p \ d^3r \ \epsilon \ \exp(-\beta \epsilon) = -\langle \epsilon \rangle \]

\[ \langle \epsilon \rangle = 3/(2\beta) = (3/2)kT \]

(The usual way)$^2$

\[ \partial^2 \ln Z / \partial \beta^2 = -\partial \langle \epsilon \rangle / \partial \beta = ??? \]

Let $N$ particles be confined to a box of volume $\Omega$. What is the RMS fluctuation in the pressure?
Later on we shall make use of a time-dependent gaussian random variable, $\eta(t)$. In its usual use, $\eta(t)$ is a very rapidly varying quantity, with a time-integral which behaves like a Gaussian random variable. Specifically, it is defined to have two properties:

\[ <\eta(t)> = 0 \]

\[ X(t) = \int_s^t du \, \eta(u) \] is a Gaussian random variable with variance $\Gamma |s-t|$. Here $\Gamma$ defines the strength of the oscillating random variable.
Approximate Gaussian Integrals

It is often necessary to calculate integrals like

\[ I = \int_a^b \, dx \, e^{Mf(x)} \]

in the limit as \( M \) goes to infinity. Then the exponential varies over a wide range and the integral appears very difficult. But, in the end it's easy. The main contribution will come at the maximum value of \( f \) in the interval \([a,b]\). Assume there is a unique maximum and the second derivative exists there. For definiteness say that the maximum occurs at \( x=0 \), with \( a<0<b \). Then we can expand the exponent and evaluate the integral as

\[ I \approx e^{Mf(0)} \int_a^b \, dx \, e^{Mf''(0)x^2/2} + \ldots = e^{Mf(0)} \left( \frac{2\pi}{-Mf''(0)} \right)^{1/2} \]

Notice that because we have assumed that zero is a maximum, the second derivative is negative. Because \( M \) is large and positive, we do not have to include any further higher order terms in \( x \). For the same reason we can extend the limits of integration to infinity. With that, it's done!

We shall have an integral just like this later on.

Let's do it now. Calculate \( I = \int dx \, [\cos x]^M \exp(ikx) \) with the integral going from 0 to \( \pi/2 \) and \( M \) being a very large positive number.
Calculation of Averages and Fluctuations: Reprise

The usual way: in general

Let's start from a Hamiltonian on a Lattice

\[ W\{\sigma\} = -\beta H\{\sigma\} = \sum_r h_r \sigma_r + \sum_{\langle r,s \rangle} K \sigma_r \sigma_s \]

This Hamiltonian defines what is called the Ising model. The first sum is a sum over all lattice sites, \( r \). The second sum is a sum over nearest neighbors. The field, \( h_r \), which depends upon \( r \), multiplies a spin variable which is different, of course, different on each site. The notation \( \{\sigma\} \) says that these things depend upon many spin variables. You can assume that the spin variables take on the value +1 or -1 if you want, but the argument is very general and the result does not depend upon what \( \sigma \) might be. Start from \( Z = \text{Tr} \exp W\{\sigma\} \) where \( \text{Tr} \) means a summation over all possible values of all the spin variables. It is a repeat of the argument that we have given before to say that

\[ <\sigma_r> = \partial \ln Z / \partial h_r \]

The partial derivative means that we hold all other \( h \)'s constant. The second derivative is
given by

\[ <\sigma_r \sigma_s> - <\sigma_r><\sigma_s> = <[\sigma_r - <\sigma_r>][\sigma_s - <\sigma_s>]> = \partial^2 \ln Z / \partial h_s \partial h_r = \partial <\sigma_s> / \partial h_r \]
To see this we go slowly

\[ \partial <\sigma_s>/ \partial h_r = \partial \frac{\text{Tr}[\sigma_s \exp(W{\sigma})]/\text{Tr} \exp(W{\sigma})}{\partial h_r} \]

This derivative can act in two places, one here and one here. In both cases the derivative bring down a factor of \( \sigma_r \). Since the denominator is called \( Z \), we have that this expression is equal to

\[ = \frac{\text{Tr}[\sigma_s \ \sigma_r \exp(W{\sigma})]/Z}{Z^2} \]

which then simplifies to

\[ = <\sigma_s \sigma_r> - <\sigma_s> <\sigma_r> \]

\[ = <[\sigma_r - <\sigma_r>][\sigma_s - <\sigma_s>]> \]

This expression is called the spin-spin correlation function. It expresses how fully the fluctuations in the values of the spins in the two places are correlated. In general this kind of correlation falls off with distance, but how it falls off tells us a lot about what is going on.
Sums and Averages in Classical Mechanics

The probability distribution for a single particle in a weakly interacting gas as is given by
\[ \rho(p, r) = \left( \frac{1}{z} \right) \exp(-\beta H) \]
\[ H = \left[ p_x^2 + p_y^2 + p_z^2 \right] / 2m + U(r) \]

Here, the potential holds the particles in a box of volume \( \Omega \), so that \( U \) is zero inside a box of this volume and infinite outside of it. The partition function, is
\[ z = \Omega \left[ \int dp \exp(-\beta p^2 / (2m)) \right]^3 = \Omega (2\pi m / \beta)^{3/2} \]
The average of any function of \( p \) and \( r \) is given by
\[ <g(p, r)> = \int d\mathbf{p} \ d\mathbf{r} \ \rho(\mathbf{p}, \mathbf{r}) \ g(\mathbf{p}, \mathbf{r}) \]

Since there are \( N \) particles in the system \( N \ d\mathbf{p} \ d\mathbf{r} \ \rho(\mathbf{p}, \mathbf{r}) \) is the number of particles which have position and momentum within \( d\mathbf{p} \ d\mathbf{r} \) about the phase space point \( \mathbf{p}, \mathbf{r} \). The quantity \( N \ \rho(\mathbf{p}, \mathbf{r})=f(\mathbf{p}, \mathbf{r}) \) is called the distribution function. The total amount of the quantity represented by \( g(\mathbf{p}, \mathbf{r}) \) is given in terms of the distribution function as
\[ \text{total amount of } g = \int d\mathbf{p} \ d\mathbf{r} \ f(\mathbf{p}, \mathbf{r}) \ g(\mathbf{p}, \mathbf{r}) \]

Example: We calculated the average energy \( < p^2 / (2m) > = 3 kT / 2 = \int d\mathbf{p} \ d\mathbf{r} \ \rho(\mathbf{p}, \mathbf{r}) \ p^2 / (2m) \)
The total energy in the system is \( \int d\mathbf{p} \ d\mathbf{r} \ f(\mathbf{p}, \mathbf{r}) \ p^2 / (2m) = 3N kT / 2. \)
More sums and averages

The normalization condition for the probability is $\int dp\, dr\, \rho(p,r) = <1> = 1$

The normalization for the distribution function is $\int dp\, dr\, f(p,r) = N$

The pressure, $P$, is defined as the total momentum transfer to a wall per unit of area and unit of time. Call these $dA$ and $dt$. Since a low density gas is the same all over, the number hitting is the number within the distance $p_x/m\, dt$ of the area, for $p_x > 0$, and hence the number within the volume $p_x/m\, dt\, dA$ which is $\int dp\, f(p,r)\, p_x/m\, dt\, dA$ with the integral covering all $p$'s with the condition that $p_x > 0$. If a particle hits the wall and bounces back it transfers momentum $2p_x$.

Therefore the total momentum transferred is $\int dp\, f(p,r)\, p_x/m\, dt\, dA\, 2p_x$ once again with the condition that $p_x > 0$. An integral over all momenta would give a result twice as large. In the end we get that the pressure is

$$P = \int dp\, f(p,r)\, p_x^2/m$$

which is then $NkT$ as we knew it would be.

The partition function is the sum over all variables of $\exp(-\beta H)$. For large $N$, it can be interpreted as $W\exp(-\beta<H>)$, where $W$ is the number of configuration which enter. Boltzmann got $W$ in terms of the entropy as $\ln W = S/k$. We put together previous results and find

$$W\exp(-3N/2) = z^N = \Omega^N (2\pi m kT)^{3N/2}$$

so that $S/k = N[\ln \Omega + 3(\ln (2\pi e m kT))/2]$
Homework

Work out the value of the pressure for a classical relativistic gas with $H = |p|c$. Do this both by using kinetic theory, as just above, and also by differentiating the partition function.

Statistical Mechanics started with Boltzmann’s calculation of $f(p,r,t)$ for a gas of $N$ particles in a non-equilibrium situation. He calculated a result for the entropy as a logarithm of the number of configurations, or in terms of an integral roughly of the form $\int dp \ f(p,r) \ln f(p,r)$. From our knowledge of equilibrium statistical mechanics how might we guess that this integral was related to the entropy?

The length of a random walk of $N$ steps is of the form of a sum of $N$ iid variables $\sigma_j$ each of which takes on the values 0 and $\pm 1$ with equal probability. What is the probability that a walker who takes 1000 steps will end up at a point at least 30 steps away from her starting point? (A relative accuracy of two decimal digits will suffice.)

The Hamiltonian for $N$ particles in a fluid is $H = \sum_{\alpha} p_\alpha^2/(2m) + \sum_{\alpha < \beta} V(r_\alpha-r_\beta)$. If the interaction, $V$, is weak we can assume that the particles move independently of one another. What is the value of the pressure?