

Section notes on statistical mechanics

McGreevy

Spring 2011

Contents

| | | |
|----------|---|-----------|
| 1 | Definitions of probability | 1 |
| 2 | Birthday problem | 2 |
| 2.1 | Bayes' rule and conditional probability | 2 |
| 2.2 | back to the birthday problem | 3 |
| 3 | Random walk in one dimension | 5 |
| 3.1 | What we've shown | 7 |
| 3.2 | Some physical realizations of the 1d random walk | 8 |
| 4 | Warmup problem: Einstein model of a solid | 9 |
| 4.1 | Thermodynamic properties of the Einstein solid | 10 |
| 5 | Expectation values | 12 |
| 5.1 | example 1: gaussian distribution | 13 |
| 5.2 | example 2: binomial distribution | 13 |
| 5.3 | Behavior of the variance under a linear change of variables | 14 |
| 6 | Pictures of binomial distribution, how to make plots in Mathematica | 14 |
| 7 | Warmup problem: two discrete random variables, conditional probabilities | 15 |
| 8 | Applications of Poisson statistics | 16 |
| 8.1 | Distance to nearest gas atom | 16 |
| 8.2 | Shot noise | 18 |
| 9 | Characteristic functions or generating functions | 19 |
| 9.1 | mean and variance of binomial distribution | 19 |
| 9.2 | Characteristic functions | 20 |

| | |
|--|-----------|
| 9.3 Behavior of the variance under a linear change of variables | 22 |
| 10 Pictures of binomial distribution, how to make plots in Mathematica | 23 |
| 11 Warmup problem: optimal selection | 23 |
| 12 An alternative to the graphical method for changing variables in probability distributions | 26 |
| 12.1 Application to pset 3 problem 4 | 26 |
| 12.2 More examples | 28 |
| 13 Comments on central limit theorem | 28 |
| 14 Comment about thermodynamic variables which are neither intensive nor extensive | 30 |
| 15 1d chain: entropic forces (another application of the binomial distribution and its gaussian peak) | 30 |
| 16 Comment on limits of integration | 32 |
| 17 Warmup problem | 32 |
| 18 Relations among response functions | 33 |
| 18.1 Comparison to 1d chain rule | 35 |
| 18.2 An even simpler example | 35 |
| 19 Exact differentials and second derivatives | 37 |
| 20 1d chain: entropic forces (another application of the binomial distribution and its gaussian peak), part 1 | 37 |
| 21 Warmup problem: exact and inexact differentials | 40 |
| 22 1d chain: entropic forces (another application of the binomial distribution and its gaussian peak) | 42 |
| 23 Adiabats and entropy: a review problem | 44 |
| 23.1 Entropy of ideal gas | 45 |
| 23.2 Entropy of VdW gas | 46 |
| 23.3 Carnot cycle in T, S plane | 47 |
| 23.4 Motivating entropy as a state variable | 48 |
| 24 Cooling a fluid: some exercises using the first law | 49 |

| | | |
|-----------|---|-----------|
| 24.1 | Joule-Thomson Throttling | 49 |
| 25 | Comments on pset 5b, problem 2: Rückhardt's experiment | 50 |
| 26 | S2008 first test, last problem. | 52 |
| 27 | A microcanonical example: Classical magnetic moments | 52 |
| 27.1 | Whence the expression for $\Omega(M, N)$? | 53 |
| 27.2 | Comment on regime of validity of the answer for $p(m)$ | 53 |
| 27.3 | Comments about microcanonical assumption for continuous classical distributions | 54 |
| 28 | Heat pumps and refrigerators | 54 |
| 29 | There is no magic about Maxwell relations | 55 |
| 30 | More paramagnet | 56 |
| 31 | Cooling a fluid: an exercise using the first law and Maxwell relations | 59 |
| 31.1 | Joule-Thomson Throttling | 59 |
| 32 | Warmup problem: defects in a solid | 61 |
| 33 | Information and entropy | 62 |
| 33.1 | Entropy and probability | 63 |
| 34 | Warmup problem: Schottky anomaly | 65 |
| 35 | Comment on derivation of canonical ensemble from microcanonical ensemble | 66 |
| 36 | Quantum ideal gas of distinguishable particles | 69 |
| 36.1 | High temperature | 69 |
| 36.2 | Aside about variance of energy | 70 |
| 36.3 | Low temperature | 71 |
| 37 | Entropy of blackbody radiation and a puzzle | 71 |
| 37.1 | Entropy of blackbody radiation | 72 |
| 37.2 | The history of the earth is consistent with thermodynamics | 73 |
| 38 | “Quietly emitting thermal radiation” ? | 73 |
| 38.1 | Hydrodynamic description | 74 |
| 38.2 | Density of states | 75 |

| | |
|--|-----------|
| 38.3 Enter the dispersion relation | 76 |
| 38.4 Periodic boundary conditions | 77 |
| 39 Warmup problem: number of photons | 78 |
| 39.1 A better proof relating entropy to occupation numbers | 78 |
| 40 Pressure of photon gas | 79 |
| 41 Adsorption: an application of chemical potential | 80 |
| 42 Another chemical potential example: liquid-vapor equilibrium | 82 |
| 43 Air | 83 |
| 43.1 Law of isothermal atmospheres from canonical ensemble | 83 |
| 43.2 Sound in air and adiabatic density variations | 84 |
| 44 Pressure of quantum gases | 85 |
| 45 Johnson-Nyquist noise of an electrical circuit at finite temperature | 87 |
| 45.1 Thermal radiation in one dimension | 87 |
| 45.2 Back to circuit noise | 88 |
| 46 Warmup question: entropy and occupation numbers | 89 |
| 47 Comment about pressure of ideal gases | 90 |
| 48 Bose condensation in two dimensions? | 91 |

1 Definitions of probability

The *probability* $P(A)$ of an event A is a real number which represents our belief about how likely that event A is to occur. If the set of possibilities (that is, A and its alternatives – this is called the *sample space*), is discrete, then $P(A)$ is between 0 and 1. For today, let’s restrict ourselves to this case. The case where there is a continuum of possibilities will be important and we’ll get to it.

The definition I’ve just described is called the ‘subjective’ (or Bayesian) definition of probability, because it makes explicit reference to our thinking. A simpler-to-make-concrete definition is called the ‘objective’ (or frequentist) definition, which is to imagine that we have a whole big collection of identically-prepared situations where A has a chance to occur (say N such situations, called *trials*), and we just count the fraction

in which A does indeed occur:

$$P(A) \stackrel{?}{=} \frac{\text{number of trials in which } A \text{ occurs}}{\text{total number of trials}} .$$

More precisely, we should make sure we do enough trials, so that the answer doesn't depend on the number of trials:

$$P(A) = \lim_{N \rightarrow \infty} \frac{\text{number of trials in which } A \text{ occurs}}{\text{total number of trials, } N} .$$

There are situations when this frequentist definition cannot be made precise, as when one is not able to prepare identical trials for one reason or another. Some practical examples include: when asking for the probability that humans are responsible for global warming, when asking for the probability that your first child will be born in the next decade, when asking about the probability of the existence of life in the universe. In this case one is forced to use the subjective definition. It is no less precise; for a beautiful and wildly entertaining discussion of the point of view that probability theory is merely the application of logic to incomplete information, I highly recommend the book E.T. Jaynes, *Probability: the Logic of Science*. Alas, we will have to retreat to the frequentist definition in this course.

A property of probability which is obvious from the frequentist definition is that the sum of the probabilities over *all* possibilities had better be one: that is, for sure *something* will happen, even if it's nothing.

$$1 = \sum_{\text{all possibilities for the value of } A} P(A) \tag{1.1}$$

Note that in (1.1) I have used the letter A as a dummy variable which runs over the whole sample space of possible events.

2 Birthday problem

The following problem will be an opportunity to practice with some important notions in probability theory.

In a group of N people, what is the probability P_N that (at least) two share a birthday?

Some simplifications: Ignore leap year. Assume all days are represented uniformly as people's birthdays. This is not quite the case. A clumpy distribution of possible birthdays increases the probability of birthday collision.

Let's take advantage of (1.1) to solve a slightly simpler equivalent problem: what's the probability \bar{P}_N that *no two* of the N people share a birthday? Since they are mutually exclusive, these two probabilities are related by

$$\bar{P}_N = 1 - P_N.$$

To figure this out let's put the people in some random order (imagine them telling you their birthday one at a time as in the first recitation), and ask for the probability that there is no collision at each step. The following digression will be useful for this.

2.1 Bayes' rule and conditional probability

We are going to need to talk about probabilities $P(x, y)$ that depend on multiple random variables, whose outcomes may or may not be correlated. In the example at hand, x can be whether person 1 and 2 share a birthday, and y can be whether person 2 and 3 share a birthday. It must be normalized so that

$$\sum_{x,y} P(x, y) = 1. \quad (2.1)$$

Note that if we sum over all possibilities for x , we get a probability distribution for y : $\sum_x P(x, y) = P(y)$, and vice versa $\sum_y P(x, y) = P(x)$, and these distributions are automatically normalized by (2.1).

Define the *conditional probability* $P(x|y)$ to be the probability that x happens, *given* that y happens. It must be proportional to $P(x, y)$ the probability that both happen: $P(x|y) = \mathcal{N}P(x, y)$. But $P(x|y)$ is a probability on the space of possible values of x . In particular this means:

$$\sum_x P(x|y) = 1$$

for each possible value of y . This gives us an expression to determine the normalization \mathcal{N} :

$$1 = \sum_x P(x|y) = \mathcal{N} \sum_x P(x, y) = \mathcal{N}P(y),$$

so $\mathcal{N} = 1/P(y)$ and we have derived 'Bayes' rule' :

$$P(x, y) = P(x|y)P(y). \quad (2.2)$$

Two outcomes x, y are said to be *statistically independent* if $P(x, y) = P(x)P(y)$ which means (from Bayes' rule) that $P(x|y) = P(x)$ and $P(y|x) = P(y)$.

Notice that the LHS of (2.2) is manifestly symmetric in the role of x, y , so

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x).$$

This means that if we want to figure out $P(x, y)$ we can do it by multiplying conditional probabilities in any order.

2.2 back to the birthday problem

Let $D = 365$ be the number of days per year. Order the people by an index $k = 1 \dots N$. The probability that no two share a birthday is

$$\bar{P}_N = \prod_{k=1 \dots N} P(\text{person } k \text{ doesn't share a birthday with any of the previously analyzed people, } 1 \dots k-1)$$

The first factor in the product is just 1, since there have been no previous people. For the second person, one possible birthday is excluded by the birthday of the first person:

$$P(\text{person 2 doesn't share a birthday with person 1}) = \frac{D-1}{D} = 1 - \frac{1}{D}.$$

For the third person, two days are excluded, and so on until the N th person. The result is:

$$\bar{P}_N = \prod_{k=1 \dots N} \left(1 - \frac{k-1}{D}\right).$$

Note that the factors in this product are conditional probabilities: we are assuming at each step that there was no collision at the previous steps; otherwise there would be fewer than $k-1$ days excluded at the k th step.

Note that the random variables associated with the collisions of birthdays are *not* independent, since they are all constrained to be taken from the same set of 365 days. (A similar situation holds for drawing cards from a deck.) If the collisions of pairs were independent (which they aren't because to specify a configuration, we only need to specify N birthdays, not $N(N-1)/2$ pairs of birthdays), we would have

$$\bar{P}_N \stackrel{?}{=} \prod_{\text{pairs, p}} P(\text{pair p don't share a birthday}) = \left(1 - \frac{1}{D}\right)^{\text{number of pairs}} = \left(1 - \frac{1}{D}\right)^{\frac{N(N-1)}{2}}.$$

This is actually not too bad an approximation if $N < D$.

So the answer to our question looks like figure ???. Note that the probability of a collision is already .5 when $N = 23$. This may be surprising if you confuse this question

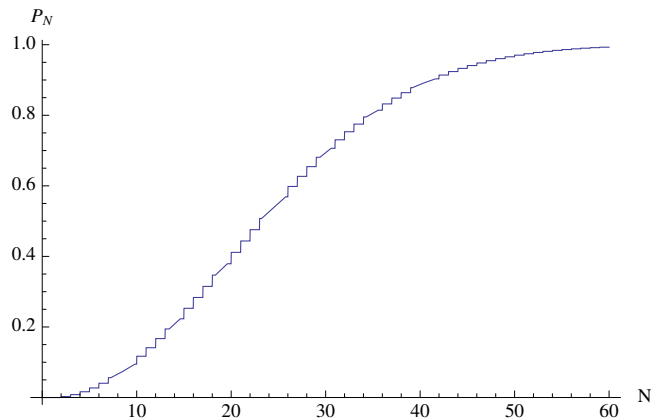


Figure 1: The probability of birthday coincidence as a function of the number of people.

with the question: what is the probability that someone in a room of N other people has the same birthday *as you*? This is much less likely:

$$P(\text{same as you}) = 1 - P(\text{none same as you}) = 1 - \left(1 - \frac{1}{D}\right)^{N-1}$$

Note that here you don't care if any of the N people share birthdays with each other, so they are independent variables. This grows much more slowly with N .

For small $x \ll 1$, $1 - x \sim e^{-x}$. By " \sim " I mean that this is a good approximation *and* an upper bound. Using this, we see (without using a computer) that the probability of no collision shrinks rapidly with N at small N :

$$\bar{P}_N \sim e^{-\sum_{k=1}^{N-1} \frac{1}{D}} = e^{-\frac{1}{D} \sum_{k=1}^{N-1} (k-1)} = e^{-\frac{N(N-1)}{2D}}.$$

A final comment about this problem. There is in fact a more direct approach to figuring out \bar{P}_N , namely to enumerate all the possible ways to assign the people birthdays and to count how many of these don't involve a collision. \bar{P}_N is the fraction without a collision. (I mention this approach only at the end of this discussion because the method above was a useful opportunity to introduce conditional probability and Bayes' rule.) The number of ways to assign birthdays is D^N , since there are D possibilities for each person's birthday. The number of these without a collision requires us to choose N distinct days. The first one we choose has D options, the second one has $D - 1$, the N th has $D - N + 1$ options. So

$$\bar{P}_N = \frac{D(D-1)\dots(D-N+1)}{D^N}$$

which agrees with our previous expression.

3 Random walk in one dimension

A drunk person is trying to get home from a bar at $x = 0$, and makes a series of steps of length L down the (one-dimensional) street. Unfortunately, the direction of each step is random, and uncorrelated with the previous steps: with probability p he goes to the right and with probability $q = 1 - p$ he goes to the left. Let's ask: after N steps, what's his probability $P(m)$ of being at $x = mL$?

Note that we've assumed all his steps are the same size, which has the effect of making space discrete. Let's restrict ourselves to the case where he moves in one dimension. This already has many physical applications, some of which we'll mention later.

What's the probability that he gets $|m| > N$ steps away? With N steps, the farthest away he can get is $|m| = N$, so for $|m| > N$, $P(m) = 0$.

Consider the probability of a particular, ordered, sequence of N steps, $x_i = L$ or R :

$$P(x_1, x_2 \dots x_N) = P(x_1)P(x_2) \cdots P(x_N) = p^{n_R}q^{n_L}.$$

In the second step here we used the fact that the steps are statistically independent, so the joint probability factorizes. n_R is the number of steps to the right, i.e. the number of the x_i which equal R . Since the total number of steps is N , $n_L + n_R = N$, the net displacement (in units of the step length L) is

$$m = n_R - n_L = 2n_R - N.$$

Note that $m = N \bmod 2$.

In asking about the drunk's probability for reaching some location, we don't care about the order of the steps. There are many more ways to end up near the starting point than close by. For example, with $N = 3$, the possibilities are

$$\begin{aligned} LLL & \quad m = -3 \\ RLL, LRL, LLR & \quad m = -1 \\ RRL, RLR, LRR & \quad m = 1 \\ RRR & \quad m = 3 \end{aligned}$$

What's the number of sequences for a given n_L, n_R ? The sequence is determined if we say which of the steps is a R, so we have to choose n_R identical objects out of N . The number of ways to do this is

$$\binom{N}{n_R} = \frac{N!}{n_R!n_L!} = \binom{N}{n_L}.$$

A way to think about this formula for the number of ways to arrange $N = n_R + n_L$ of which n_R are indistinguishably one type and n_L are indistinguishably another type, is: $N!$ is the total number of orderings if all the objects can be distinguished. Redistributing the n_R R -steps amongst themselves doesn't change the pattern (there are $n_R!$ such orderings), so we must divide by this overcounting. Similarly redistributing the n_L L -steps amongst themselves doesn't change the pattern (there are $n_L!$ such orderings).

So

$$P(n_L, n_R) = \frac{N!}{n_R!n_L!} p^{n_R} q^{n_L}.$$

Note that the binomial formula is

$$(p + q)^N = \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n}.$$

Since we have $p + q = 1$, this tells us that our probability distribution is normalized:

$$\sum_{n_R=0}^N \frac{N!}{n_R!(N-n_R)!} p^{n_R} q^{N-n_R} = 1^N = 1.$$

The probability for net displacement m is

$$P(m) = \frac{N!}{\left(\frac{N+m}{2}\right)! \left(\frac{N-m}{2}\right)!} p^{\frac{N+m}{2}} q^{\frac{N-m}{2}}$$

for $N \pm m$ even, and zero otherwise.

3.1 What we've shown

We have shown that the probability that an event with probability p occurs n times in N (independent) trials is

$$W_N(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} ;$$

this is called the binomial distribution. $1-p$ here is the probability that *anything* else happens. So the analog of "step to the right" could be "a particular song is played on your ipod in shuffle mode" and the analog of "step to the left" is "any other song comes on".

For example, suppose you have 2000 songs on your ipod and you listen on shuffle by song; then the probability of hearing any one song is $p = \frac{1}{2000}$. Q: If you listen to

$N = 1000$ songs on shuffle, what's the probability that you hear a particular song n times?

The binomial distribution applies. But there are some simplifications we can make. First, p itself is a small number, and N is large. Second, the probability will obviously be very small for $n \sim N$, so let's consider the limit $n \ll N$. In this case, we can apply Sterling's formula to the factorials:

$$W_N(n) \approx \frac{1}{n!} \frac{N^N}{(N-n)^{N-n}} p^n (1-p)^{N-n}$$

We can use $N - n \sim N$ except when there is a cancellation of order- N terms:

$$W_N(n) \approx \frac{1}{n!} \frac{N^N}{(N)^{N-n}} p^n (1-p)^{N-n} = \frac{1}{n!} N^n p^n (1-p)^{N-n}$$

Now we can Taylor expand in small p , using $\ln(1-x) \approx -x + x^2/2 - x^3/3 + \dots$

$$W_N(n) \approx \frac{1}{n!} (Np)^n e^{(N-n)\ln(1-p)} \approx \frac{1}{n!} (Np)^n e^{-Np}.$$

This is called the Poisson distribution,

$$\text{Poisson}_\mu(n) = \frac{1}{n!} \mu^n e^{-\mu}.$$

Note that it only depends on the product $\mu = pN$, which for our example is $pN = \frac{1}{2000} 1000 = 1/2$. In this case, it looks like in the figure ??.

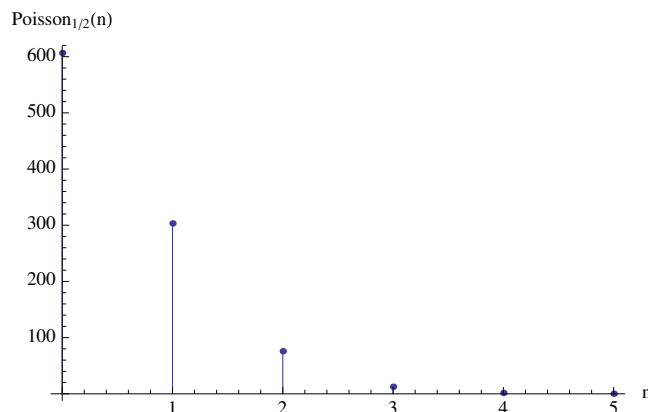


Figure 2: The Poisson distribution for $pN = 1/2$, $\text{Poisson}_{1/2}(n)$.

It may seem like your ipod is conspiring to play some songs multiple times and not play others at all (I had this impression too until I thought about it), but it's just because we don't have much intuition yet about the Poisson distribution. In fact, if we vary $\mu = Np$, we can make the probability that a given song is never heard much larger than the probability that it is heard once; see figure ??.

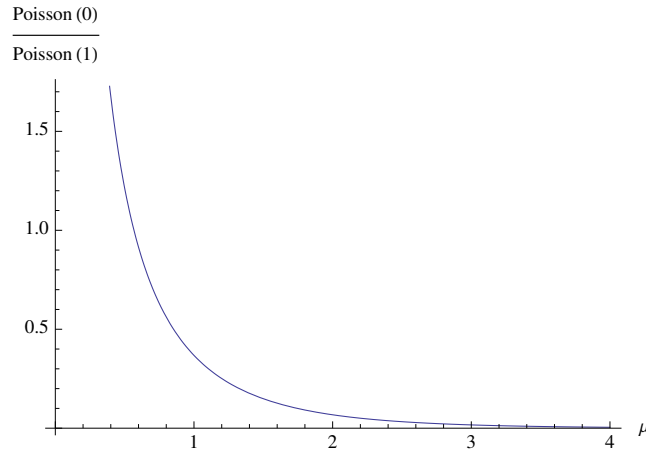


Figure 3: The ratio of the poisson distribution at $n = 0$ to $n = 1$ as we vary the parameter μ . (Note that this figure is not a probability distribution.)

3.2 Some physical realizations of the 1d random walk

Besides modelling flipping of coins, and shuffle of mp3s, this is also a useful model for several physical systems.

One example is spins in a magnet. The magnetic moment of an atom with spin $1/2$ can point up or down. Sometimes this is random. What's the net magnetic moment of N such atoms?

Another example is the following. Consider N gas molecules in a container of volume V . Let ν be the volume of some specific small subregion of the container. Let n be the number of molecules in this small subregion. What's the distribution $P(n)$?

Let $n_{i=1..N}$ be a function of the positions of the particles, which is 1 if the i th particle is in the subregion of interest, and 0 otherwise. So $n = \sum_{i=1}^N n_i$. Assume two things:

- (1) n_i are independent (particles don't interact or are not otherwise correlated).
- (2) the probability of $n_i = 1$ is $p = \nu/V$, *i.e.* each particle has equal probability of being anywhere in the big volume.

Then this is another situation where our binomial distribution applies:

$$P(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

In the limit where N is large, and V is large compared to ν , we can again do the manipulations that led to the Poisson distribution:

$$P(n) \approx \frac{(\alpha\nu)^n}{n!} e^{-\alpha\nu}$$

where $\alpha = N/V$ is the density of particles.

4 Warmup problem: Einstein model of a solid

[from *e.g.* Reichl, *A Modern Course in Statistical Physics*]

Here's a combinatorics problem which comes up in Einstein's model of a solid; this is a name for the model discussed in pset 2 problem 1 (warning: this problem is NOT the same as the question asked there).

Q: Given a number q of *identical* balls, how many ways are there to put them into N *distinct* bins? Call the answer $\mathcal{N}_N(q)$.

A: Line up the bins; denote the boundaries of the bins between the balls by vertical lines. Here is one arrangement

$$\circ | \circ || \circ \circ |$$

which has $q = 4$ balls and $N = 5$ bins. Here is another with the same q, N :

$$\circ | \circ ||| \circ \circ .$$

To specify such an arrangement we just need to say where to put the barriers. The number of ways to do this is just like in the 1d random walk with a fixed number of steps: the number of ways to arrange $q + N - 1$ objects where $N - 1$ of them are of one type and q of them are of another type is

$$\mathcal{N}_N(q) = \binom{q + N - 1}{q} = \frac{(q + N - 1)!}{q!(N - 1)!}.$$

The physical application I mentioned involves some quantum mechanics. Consider a collection of N harmonic oscillators (masses attached to springs), all with natural frequency ω which are arrayed in space somehow. They are distinguishable by their location. (This is the *Einstein* model of a solid because the oscillators don't interact with each other; interactions were included by Debye.) The possible states of each oscillator is labelled by a whole number $m_i = 0, 1, 2, \dots$ indicating how excited it is; this number is called a "number of quanta". The total energy is determined by the sum of these numbers (this is where it's important that the oscillators are harmonic) $q = \sum_{i=1}^N m_i$, the total number of quanta. So if we think of m_i as the number of balls in bin i , the number of ways to distribute these quanta among the oscillators is $\mathcal{N}_N(q)$; this is the number of microstates of the Einstein solid with fixed energy.

We can come back to this later and use this result to compute thermodynamic properties of the resulting solid.

4.1 Thermodynamic properties of the Einstein solid

We can use the preceding result to compute thermodynamic properties of the resulting solid. Recall the relation between entropy and temperature:

$$dS = \frac{dE}{T}.$$

On the LHS dS is the change in entropy of a system to which we just added heat at temperature T . On the RHS dE is the amount of heat we added, *i.e.* by how much the internal energy of the system increased.

We've just figured out $S_N(E)$ for the Einstein solid, because we know $S(q)$ and we know E in terms of q . We have

$$S_N(q) = k_B \ln \left(\frac{(N-1+q)!}{(N-1)!q!} \right) \approx k_B \ln \left(\frac{(N+q)^{N+q}}{N^N q^q} \right) = k_B ((N+q) \ln(N+q) - q \ln q - N \ln N)$$

where we used $N \gg 1$ and Stirling. But as we said above the number of quanta q determines the energy; the relevant relation is

$$E(q) = \sum_{i=1}^N \hbar\omega(m_i + \frac{1}{2}) = \hbar\omega(q + N/2).$$

So the above formula gives $S_N(E)$.

Using the key relation between temperature and entropy, we can find the energy of the solid as a function of the temperature:

$$\frac{1}{T} = \frac{dS}{dE} = \frac{dq}{dE} \frac{dS}{dq} = \frac{1}{\hbar\omega} k_B (+\ln(N+q) + 1 - \ln q - 1)$$

which says that

$$\frac{\hbar\omega}{k_B T} = \ln \left(\frac{N+q}{q} \right).$$

Notice that the LHS is dimensionless (as it should be since the LHS is) because both $k_B T$ and $\hbar\omega$ have dimensions of energy. Let's exponentiate the both hand side (BHS) and solve for q :

$$q = \frac{N e^{-\frac{\hbar\omega}{k_B T}}}{1 - e^{-\frac{\hbar\omega}{k_B T}}}$$

and hence

$$E = N \frac{1}{2} \hbar\omega + N \hbar\omega \frac{e^{-\frac{\hbar\omega}{k_B T}}}{1 - e^{-\frac{\hbar\omega}{k_B T}}}$$

The first term is the sum of the zero-point energies of all the oscillators. The second term depends on the temperature. You will soon recognize the structure of this function as characteristic of a system of bosons in thermal equilibrium.

This formula says how much energy you have to put in to raise the temperature of the solid, which is something that's quite measurable; it gives the right answer if the temperature isn't too low. (For lower temperatures one must take into account the interactions between the oscillators; this is the Debye model.)

5 Expectation values

[Blundell ×2]

Given a probability distribution for a random variable x we can define its *average value* or *mean* or *expectation value*:

$$\langle x \rangle \equiv \sum_x P(x)x$$

(we can do this as long as x takes values in a set with a rule for addition!). Sometimes this is denoted \bar{x} . Note that the mean value of x is not necessarily a value that it can take within the distribution. For example, the average number of children per family is 2.4.

We can also define *moments* of the distribution,

$$\langle x^n \rangle \equiv \sum_x P(x)x^n$$

(assuming now the variable lives in a space with a multiplication rule). If we know enough of these, the distribution P is specified. Indeed, we can define the expectation value of any function of the random variable:

$$\langle f(x) \rangle \equiv \sum_x P(x)f(x).$$

Notice that taking expectation values is a linear operation:

$$\langle af(x) + bg(x) \rangle = a \langle f(x) \rangle + b \langle g(x) \rangle$$

if a, b are constants independent of x .

Suppose we want to know how well the mean value of x represents what will happen, *e.g.* how certain are we that the value we get for x will be $\langle x \rangle$. Consider the variable

$x - \langle x \rangle$ which represents the deviation from the mean value. Its own expectation value is

$$\langle x - \langle x \rangle \rangle = 0$$

so this isn't so useful for telling us about the distribution. But the mean of its square is

$$\sigma_x^2 \equiv \langle (x - \langle x \rangle)^2 \rangle$$

called the variance. Its square root is the standard deviation: $\sigma_x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$.

5.1 example 1: gaussian distribution

At this point I must admit that we could be talking about a continuous distribution. In that case $p(x)dx$ represents the probability to find x in the interval $(x, x + dx)$, and the summation symbols I've been writing as \sum_x really represent integrals, $\int dx$.

To get a feel for the meaning of σ_x , consider the distribution

$$p(x) = Ce^{-(x-a)^2/2s^2}$$

on the real line, $x \in (-\infty, \infty)$.

Determine C so that $\int_{-\infty}^{\infty} dx p(x) = 1$.

Show that $\langle x \rangle = a$.

Show that $\sigma_x = s$.

Please see Tom Greytak's probability notes, page 13 for how to do the integrals.

The gaussian distribution is special. It is the only distribution on the real line whose higher cumulants (the moments of the deviation $\langle (x - \langle x \rangle)^n \rangle$ with $n > 2$) all vanish. For reasons that will hopefully become clear, this makes it ubiquitous in situations with many random variables; this ubiquity is called the Central Limit Theorem.

5.2 example 2: binomial distribution

The distribution for the number of steps to the left n_L after N steps of a 1d random walk is

$$W_N(n_L) = \frac{N!}{n_L!(N - n_L)!} p^{n_L} q^{N - n_L}$$

(this is for $0 \leq n_L \leq N$, otherwise the probability is zero). What is $\langle n_L \rangle$?

$$\langle n_L \rangle = \sum_{n_L=0}^N \frac{N!}{n_L!(N - n_L)!} q^{N - n_L} p^{n_L} n_L.$$

This would be a binomial sum if not for the annoying factor of n_L in there. Here is a wonderful math trick: Observe that $n_L p^{n_L} = p \partial_p p^{n_L}$. Suppose temporarily that p and q were independent, and not related by $p + q = 1$. Then the dependence on p appears only in this p^{n_L} and we could write:

$$\begin{aligned}\langle n_L \rangle &= p \partial_p \frac{N!}{n_L! (N - n_L)!} q^{N - n_L} p^{n_L} \\ &= p \partial_p (p + q)^N \\ &= N p (p + q)^{N - 1}.\end{aligned}\tag{5.1}$$

Our journey to this imaginary¹ world where probabilities don't sum to one has allowed us to do the sum; now we restore unitarity by setting $p + q = 1$ and returning home with our answer:

$$\langle n_L \rangle = pN.$$

In retrospect this answer is obvious from the definition of p as the probability of a left step

$$p = \frac{\text{number of left steps}}{\text{total number of steps}} = \frac{\langle n_L \rangle}{N}.$$

Similarly, $\langle n_R \rangle = Nq$, and $\langle n_R + n_L \rangle = Nq + Np = N(q + p) = N$. The average displacement is

$$\langle m \rangle \equiv \langle n_L - n_R \rangle = N(p - q).$$

Notice that this vanishes if $p = q$.

A fun further example is to work out the variance of the distribution for n_L using this trick of differentiating under the sum. You should do this. Rather than doing this one example here, let's discuss a wider class of examples that exhibit the same important behavior as a function of N .

5.3 Behavior of the variance under a linear change of variables

Suppose x is a random variable valued on the real line, and $y = ax + b$.

Q1: Relate $\langle y \rangle$ and $\langle x \rangle$:

$$\langle y \rangle = a \langle x \rangle + b,$$

since taking expectation values is a linear operation.

Q2: Show that

$$\sigma_y = a \sigma_x.$$

¹Actually it's not so hard to introduce simple modifications of the model which make this not so crazy-seeming: e.g. you could allow a third possibility at each step (say the walker just stays where he is) with infinitesimal probability). But how one does this doesn't affect the answer.

6 Pictures of binomial distribution, how to make plots in Mathematica

The discussion the previous set of notes about approximating the binomial distribution as a Poisson distribution was valid when two things happen: when N (the number of trials) is large, and when p (the probability of our event (e.g. left step) on each trial) is small. If N is large but p is not small, the distribution instead approaches a gaussian distribution. I won't give a slick way to prove this analytically right now, but it is vividly demonstrated by just looking at some plots.

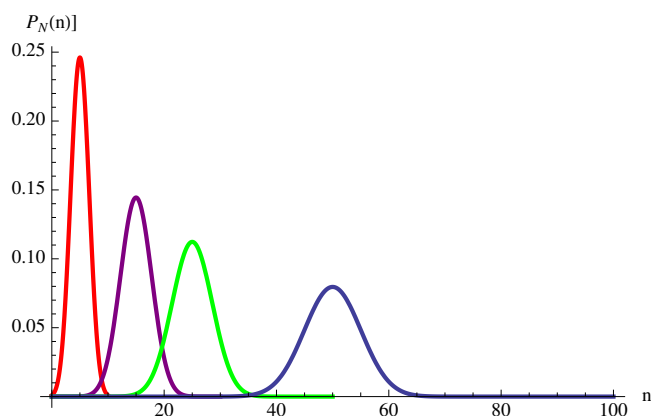


Figure 4: Binomial distributions $P_N(n)$ for various $N = 10, 30, 50, 100$ in red, purple, green blue respectively.

The following command will make one of the plots shown above:

```
NN=100;  
Plot[ Factorial[NN]/(Factorial[n]Factorial[NN-n]) (1/2)^n (1/2)^(NN-n),{n,0,NN}];  
Clear[NN];
```

7 Warmup problem: two discrete random variables, conditional probabilities

[from 8.044 S2008 pset 2]

Consider the distribution on two integer random variables

$$p(n, l) = \begin{cases} ce^{-an}, n \geq 0, & |l| \leq n \\ 0, & \text{otherwise.} \end{cases}$$

a is a constant constrained by $e^{-a} < 1$; c is a constant determined by normalizing the distribution (it depends on a).

(a) Find $p(n)$

$$p(n) = \sum_{l=-n}^n p(n, l) = (2n + 1)ce^{-an}$$

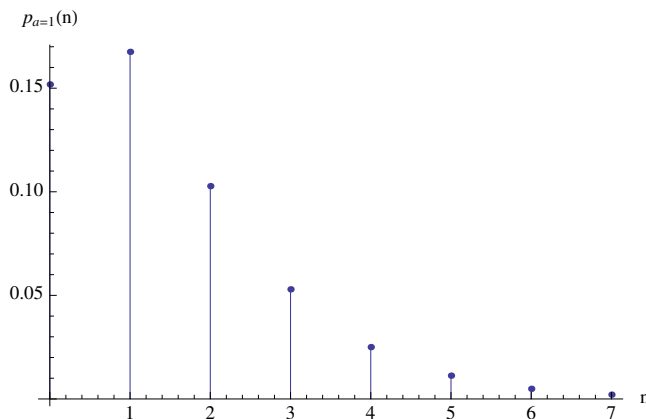


Figure 5: Part (a): $p(n)$ for the case $a = 1$. The normalization constant c is rather horrible.

(b) Find $p(l|n)$

$$p(l|n) = \frac{p(n, l)}{p(n)} = \frac{1}{2n + 1}$$

It's just flat, the uniform distribution over the range allowed by the given value of n .

(c) Find $p(l)$

$$p(l) = \sum_{n \geq |l|}^{\infty} ce^{-an} = \sum_{m \equiv n - |l|=0}^{\infty} e^{-am} ce^{-a|l|} = \frac{c}{1 - e^{-a}} e^{-a|l|}.$$

(d) Find $p(n|l)$.

$$p(n|l) = \frac{p(n, l)}{p(l)} = e^{-a(n-|l|)}(1 - e^{-a}).$$

8 Applications of Poisson statistics

8.1 Distance to nearest gas atom

Recall that in a gas of atoms with random uncorrelated positions, the number of gas molecules in a subvolume satisfies Poisson statistics. Consider N gas molecules in a container of volume V . Then

$$P(\text{one gas atom in infinitesimal volume } dv) = \frac{N}{V} dv$$

(and the probability of more than one is negligible). Let v be the volume of some specific subregion of the container. Then we showed previously that

$$P(m \text{ atoms in volume } v) = e^{-Nv/V} \frac{(Nv/V)^m}{m!} \equiv \text{Poisson}_{Nv/V}(m).$$

2

What's the distribution of distances between neighboring gas atoms, $P(r) \equiv P(\text{distance nearest gas atom} > r)$?

$$\begin{aligned} P(r)dr &= P(\text{no atom within radius } r) \cdot P(1 \text{ atom between } r \text{ and } r + dr) \\ &= \text{Poisson}_{Nv/V}(0)|_{v=\frac{4}{3}\pi r^3} \cdot \text{Poisson}_{Ndv/V}(1)|_{dv=4\pi r^2 dr} \\ &= e^{-\frac{4}{3}\pi r^3 N/V} 4\pi r^2 dr e^{-4\pi r^2 dr N/V} = e^{-\frac{4}{3}\pi r^3 N/V} 4\pi r^2 dr. \end{aligned}$$

²Here's how we did it: Let $m_{i=1..N}$ be a function of the positions of the particles, which is 1 if the i th particle is in the subregion of interest, and 0 otherwise. So $m = \sum_{i=1}^N m_i$. Assume two things:
(1) m_i are independent (particles don't interact or are not otherwise correlated).
(2) the probability of $m_i = 1$ is $p = \nu/V$, *i.e.* each particle has equal probability of being anywhere in the big volume.

Then this is another situation where our binomial distribution applies:

$$P(m) = \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m}$$

In the limit where N is large, and V is large compared to ν , we can again do the manipulations that led to the Poisson distribution:

$$P(m) \approx \frac{(\alpha\nu)^m}{m!} e^{-\alpha\nu}$$

where $\alpha = N/V$ is the density of particles.

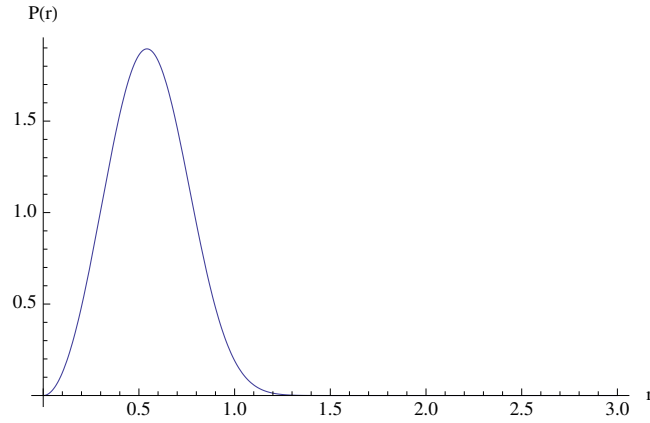


Figure 6: The distribution of distances to the nearest gas atom, for $N/V = 1/\text{length unit}^3$ in whatever length unit you like.

The max probability occurs at

$$0 = \partial_r \left(-\frac{4}{3}\pi r^3 N/V + \ln(4\pi r^2) \right) = -4\pi r^2 N/V + 2/R$$

that is

$$r_{\max} = \left(\frac{V}{2\pi N} \right)^{1/3}$$

which, up to numerical factors, should not be surprising.

8.2 Shot noise

[from Balian]

Consider a vacuum tube, aka ‘thermionic emitter’ which is a hot metal plate spitting out electrons across a vacuum to produce a current I , which on average is

$$\langle I \rangle = e\nu$$

where ν is the average number of electrons emitted per unit time and $e = 1.6 \times 10^{-19} \text{C}$ is the charge of an electron. The discrete nature of the charge carriers leads to fluctuations in the current $\Delta I \equiv \sqrt{\langle I^2 \rangle - \langle I \rangle^2}$. These fluctuations can be used to measure the charge of the electron. Suppose we measure I for one second at a current of $1 \mu\text{A}$.

(a) What’s $P(n \text{ electrons emitted in time } t)$?

Since the probability for one emission in interval dt is νdt , if we assume the emission events are independent, this satisfies the criteria for Poisson:

$$P(n \text{ electrons emitted in time } t) = \text{Poisson}_{\nu t}(n) = e^{-\nu t} \frac{(\nu t)^n}{n!}.$$

(b) What’s ΔI ?

The current we measure if we average over a time t ($= 1 \text{ sec}$ in the example) is

$$\begin{aligned} I &= \frac{(\text{number of electrons, } n)e}{t} \\ \langle I \rangle &= \frac{\langle n \rangle e}{t} = \frac{\nu t e}{t} = \nu e. \\ \langle I^2 \rangle &= \frac{\langle n^2 \rangle e^2}{t^2} = \frac{((\nu t)^2 + \nu t) e^2}{t^2} \\ \Delta I^2 &= \langle I^2 \rangle - \langle I \rangle^2 = \frac{\nu t e^2}{t^2} = \frac{\nu e^2}{t}. \end{aligned}$$

Notice that the dependence of ΔI on e, ν is different from that of I and so if we know both I and ΔI , we can figure out e .

Plugging in numbers: if $\langle I \rangle = 1 \mu\text{A}$ then $\nu = \langle I \rangle / e = 10^{13} e^- / \text{s}$ and

$$\Delta I = e \sqrt{\frac{\nu}{t}} \sim 10^{-12.5} \text{A}$$

which is apparently measurable by amplifying the resulting potential difference across a big resistor.

9 Characteristic functions or generating functions

9.1 mean and variance of binomial distribution

The distribution for the number of steps to the left n_L after N steps of a 1d random walk is

$$W_N(n_L) = \frac{N!}{n_L!(N - n_L)!} p^{n_L} q^{N - n_L}$$

(this is for $0 \leq n_L \leq N$, otherwise the probability is zero). What is $\langle n_L \rangle$?

$$\langle n_L \rangle = \sum_{n_L=0}^N \frac{N!}{n_L!(N - n_L)!} q^{N - n_L} p^{n_L} n_L.$$

This would be a binomial sum if not for the annoying factor of n_L in there. Here is a wonderful math trick: Observe that $n_L p^{n_L} = p \partial_p p^{n_L}$. Suppose temporarily that p and q were independent, and not related by $p + q = 1$. Then the dependence on p appears only in this p^{n_L} and we could write:

$$\begin{aligned} \langle n_L \rangle &= p \partial_p \frac{N!}{n_L!(N - n_L)!} q^{N - n_L} p^{n_L} \\ &= p \partial_p (p + q)^N \\ &= N p (p + q)^{N - 1}. \end{aligned} \tag{9.1}$$

Our journey to this imaginary³ world where probabilities don't sum to one has allowed us to do the sum; now we restore unitarity by setting $p + q = 1$ and returning home with our answer:

$$\langle n_L \rangle = pN.$$

In retrospect this answer is obvious from the definition of p as the probability of a left step

$$p = \frac{\text{number of left steps}}{\text{total number of steps}} = \frac{\langle n_L \rangle}{N}.$$

Similarly, $\langle n_R \rangle = Nq$, and $\langle n_R + n_L \rangle = Nq + Np = N(q + p) = N$. The average displacement is

$$\langle m \rangle \equiv \langle n_L - n_R \rangle = N(p - q).$$

Notice that this vanishes if $p = q$.

A fun further example is to work out the variance of the distribution for n_L using this trick of differentiating under the sum. You should do this. Rather than doing this

³Actually it's not so hard to introduce simple modifications of the model which make this not so crazy-seeming: e.g. you could allow a third possibility at each step (say the walker just stays where he is) with infinitesimal probability). But how one does this doesn't affect the answer.

one example here, let's discuss a machine for computing and packaging the information about the moments of a distribution.

9.2 Characteristic functions

Recall moments:

$$\langle x^m \rangle \equiv \sum_x p(x) x^m = \int dx p(x) x^m.$$

A useful device for packaging the moments of a distribution is called its *characteristic function*:

$$\tilde{p}(k) = \langle e^{-ikx} \rangle = \int dx p(x) e^{-ikx} = \left\langle \sum_{m=0}^{\infty} \frac{(-ik)^m}{m!} x^m \right\rangle = \sum_{m=0}^{\infty} \frac{(-ik)^m}{m!} \langle x^m \rangle.$$

It's the Fourier transform of the function $p(x)$. Note that it is a generating function for the moments:

$$(\partial_{-ik})^m \tilde{p}(k)|_{k=0} = \langle x^m \rangle.$$

The distribution can be returned by inverse fourier transform:

$$p(x) = \int \frac{dk}{2\pi} \tilde{p}(k) e^{+ikx}.$$

The generating function for the cumulants $\langle x^m \rangle_c$ of the distribution is the log of $\tilde{p}(k)$:

$$\ln \tilde{p}(k) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c$$

The first two cumulants are familiar, the mean and variance:

$$\langle x \rangle_c = \langle x \rangle \quad \langle x^2 \rangle_c = \sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2.$$

You can figure out the relationships between cumulants and moments by comparing the taylor coefficients of $\ln f(k)$ with those of $f(k)$.

Some examples.

1. Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-\lambda)^2}{2s^2}}$$

$$\tilde{p}(k) = \int dx \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-\lambda)^2}{2s^2} - ikx} = e^{-ik\lambda - k^2 s^2 / 2}.$$

2. Binomial distribution

$$P_N(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}$$

$$\tilde{p}_N(k) = \langle e^{-ikn} \rangle = \sum_{n=0}^{\infty} e^{-ikn} \frac{N!}{n!(N-n)!} p^n q^{N-n} \sum_{n=0}^{\infty} \frac{N!}{n!(N-n)!} (e^{-ik} p q^{-1})^n q^N = (pe^{-ik} + q)^N$$

Notice that the characteristic function *factorizes*:

$$\ln \tilde{p}_N(k) = N \ln(pe^{-ik} + q) = N \ln \tilde{p}_1(k)$$

\tilde{p} is a product of N copies of the characteristic function for one trial (and hence its log is a sum of N copies). This is a key property of independent random variables.

3. Poisson distribution

$$\tilde{p}(k) = \sum_{n=0}^{\infty} \frac{1}{n!} \mu^n e^{-\mu} e^{-ikn} = e^{-\mu} \sum_n \frac{1}{n!} (\mu e^{-ik})^n = e^{-\mu} e^{\mu e^{-ik}} = e^{\mu(e^{-ik} - 1)}.$$

$$\ln \tilde{p}(k) = \mu (e^{-ik} - 1) = \mu (-ik + (-ik)^2/2 + \dots)$$

All cumulants are the same and they are all equal to μ .

4. Exponential distribution

$$p(x) = A e^{-x/\lambda}$$

Normalization demands that $A = 1/\lambda$.

$$\tilde{p}(k) = \frac{1}{1 + ik\lambda}$$

This construction also works and is useful for multiple random variables, *e.g.* given $p(x_1, x_2, \dots, x_N)$, we can construct

$$\tilde{p}(k_1, k_2, \dots, k_N) \equiv \langle e^{-ik_1 x_1 - ik_2 x_2 \dots} \rangle.$$

If we are interested *e.g.* in just the sum of these random variables, $X \equiv \sum_i^N x_i$, then we can focus on

$$\tilde{p}(k_1 = k_2 = \dots = k_N = k) = \left\langle e^{-ik \sum_{j=1}^N x_j} \right\rangle.$$

Notice that this is the characteristic function for (*i.e.* the fourier transform of) the reduced probability distribution for X (the squashed mountain) $p(X)$:

$$p(X) = \int dx_1 \dots dx_N p(x_1, \dots, x_N) \delta(X - \sum_j x_j).$$

Hint: I'm telling you all this stuff in this section because of problem 1 on pset 3. A very useful intermediate step there is to find

$$\tilde{p}(k) = \langle e^{-ik(n_1+n_2+n_3)} \rangle$$

where $n_{1,2,3}$ are the results of the three dice throws.

9.3 Behavior of the variance under a linear change of variables

Suppose x is a random variable valued on the real line, and $y = ax + b$.

Q1: Relate $\langle y \rangle$ and $\langle x \rangle$:

$$\langle y \rangle = a \langle x \rangle + b,$$

since taking expectation values is a linear operation.

Q2: Show that

$$\sigma_y = a\sigma_x.$$

10 Pictures of binomial distribution, how to make plots in Mathematica

The discussion the previous set of notes about approximating the binomial distribution as a Poisson distribution was valid when two things happen: when N (the number of trials) is large, and when p (the probability of our event (e.g. left step) on each trial) is small. If N is large but p is not small, the distribution instead approaches a gaussian distribution. I won't give a slick way to prove this analytically right now, but it is vividly demonstrated by just looking at some plots.

The following command will make one of the plots shown above:

⁴One way to check this is to use the Fourier representation of the delta function:

$$\delta(X - \sum_j x_j) = \int \frac{dk}{2\pi} e^{ik(X - \sum_j x_j)}.$$

Alternatively, just plug in the definitions:

$$\int dX p(X) e^{-ikX} = \int dX \int \prod dx_j p(x_1 \dots x_N) \delta(X - \sum_j x_j) e^{-ikX} = \langle e^{-ik \sum x_j} \rangle$$

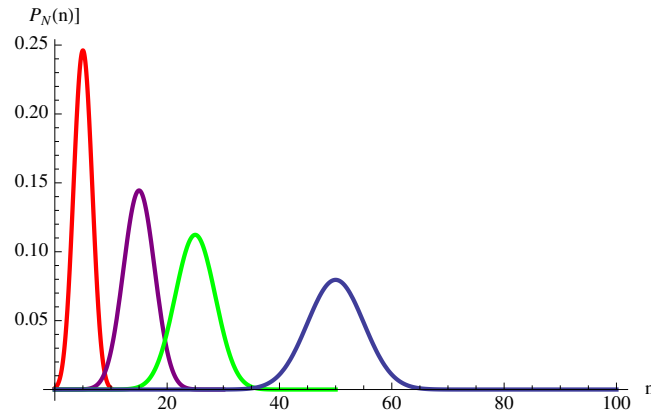


Figure 7: Binomial distributions $P_N(n)$ for various $N = 10, 30, 50, 100$ in red, purple, green blue respectively.

```

NN=100;
Plot[ Factorial[NN]/(Factorial[n]Factorial[NN-n]) (1/2)^n (1/2)^(NN-n), {n,0,NN}];
Clear[NN];

```

11 Warmup problem: optimal selection

[from Kardar, *Statistical Physics of Particles*, problem 2.4]

“In many specialized populations, there is little variability among the members. Is this a natural consequence of optimal selection?”

For example, take a gazillion random humans and break them into groups of n . Make each group of n race some fixed distance, say 26.2 miles, and record their average speed. Normalize these results by the world record, which is (amazingly) $\frac{26.2 \text{ miles}}{2:03:59 = 123.98 \text{ minutes}}$ (held by Haile Gebrselassie since 2008), so $r = \frac{\text{speed}}{\text{world record speed}} \in [0, 1]$. (For people who don’t make it the whole distance (*e.g.* are infants and can’t walk yet) the result is $r = 0$.) Now suppose we select the largest result from each group of n , *i.e.* the winner of each race, and call it x . What will be the distribution of values of x ? If we pick n big enough, so that we find an elite marathon runner in each group, they are going to cluster pretty close together.

(a) Let $\{r_\alpha\}$ be n random numbers, each independently chosen from a probability density $p(r)$, with $r \in [0, 1]$. Find the probability density $p_n(x)$ for the *largest* value in the set, *i.e.* for $x = \max\{r_1, \dots, r_n\}$.

This is a great application of the graphical method for changing variables.

A Find the region R_η of $r_1 \dots r_N$ where the max is less than some value η : $R_\eta = \{r_1 \dots r_n | x \leq \eta\}$. This region is simple to understand: if the *biggest* of the r_α is less than η , then *all* of them must be less than η !

$$R_\eta = [0, \eta] \times [0, \eta] \times \dots \times [0, \eta] = [0, \eta]^n.$$

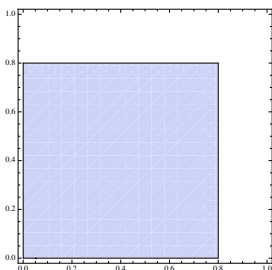


Figure 8: The region where the maximum of the random variables is less than $\eta = .8$ for the case $n = 2$.

B Find an integral expression for the cumulative probability:

$$P_x(\eta) = \int_{R_\eta} dr_1 \dots dr_n p(r_1) \dots p(r_n) = \int_0^\eta dr_1 p(r_1) \dots \int_0^\eta p(r_n) = \left(\int_0^\eta dr p(r) \right)^n$$

C Differentiate:

$$p_x(\eta) = \partial_\eta P_x(\eta) = \partial_\eta \left(\int_0^\eta dr p(r) \right)^n = n p(\eta) \left(\int_0^\eta dr p(r) \right)^{n-1}.$$

(b) Consider the special case where $p(r)$ is constant. Find $\langle x \rangle$, $\text{Var}(x)$ for this case and think about what happens at large n .

$$\text{If } p(r) = \begin{cases} 1, & r \in [0, 1] \\ 0, & \text{else} \end{cases} \quad \text{then } \int_0^\eta dr p(r) = \eta \text{ and}$$

$$p(x) = n p(x) x^n = \begin{cases} n x^{n-1}, & x \in [0, 1] \\ 0, & \text{else} \end{cases}$$

Note that the distribution is properly normalized $1 = \int_0^1 p(x) dx = \int_0^1 n x^{n-1} dx$.

So

$$\langle x \rangle = \int_0^1 dx n x^{n-1} = \frac{n x^n}{n+1} \Big|_{x=0}^1 = \frac{n}{n+1} \stackrel{n \rightarrow \infty}{\approx} 1$$

$$\langle x^2 \rangle = \int_0^1 dx n x^{n+1} = \frac{n x^{n+2}}{n+2} \Big|_{x=0}^1 = \frac{n}{n+2}$$

$$\text{Var}x = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n(n+1)^2 - (n+2)n}{(n+1)^2(n+2)} = \frac{n}{(n+1)^2(n+2)} \stackrel{n \rightarrow \infty}{\approx} \frac{1}{n^2}$$

At large n , the mean approaches 1 and the variance dies rapidly. The answer to the question is ‘yes’, at least under the given assumptions.

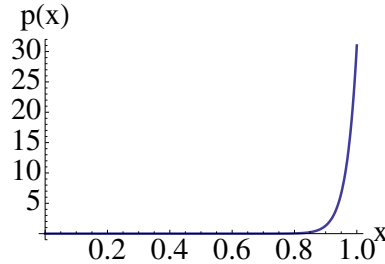


Figure 9: The distribution for the max of $n = 31$ random variables uniformly distributed over the unit interval. It is very sharply peaked around the upper limit!

Note that if we were asked about the distribution for *minimum* in each set of n , we wouldn’t have had as easy a time using this graphical method. But there’s an easy modification of it: instead of computing the cumulative probability $P(\eta) = \int_{R_\eta} p(r_\alpha) dr_1 \dots dr_n$ we could compute

$$1 - P(\eta) = \int_{\text{region where } \min\{r_1 \dots r_n\} = \eta} p(r_\alpha) dr_1 \dots dr_n$$

and then find $p(x)$ by

$$p(x) = -\partial_x(1 - P(x)).$$

12 An alternative to the graphical method for changing variables in probability distributions

Despite the previous ringing endorsement of the graphical method we have learned, I want to say some more words about changing variables in probability distributions. I will describe it in the context of Problem set 3 number 7e. This is an example where we have a distribution for some random variables, and we want to find the resulting distribution for a function of those variables. So you could use the ABC procedure that Krishna described in lecture and which is in Greytak’s notes, and which is exemplified above. Alternatively, there may be cases where you might prefer the more succinct prescription:

$$p_{x_1-x_2}(\alpha) = \int_{\text{all values}} dx_1 dx_2 p(x_1, x_2) \delta(\alpha - (x_1 - x_2))$$

where δ is the dirac delta function, which will set the dummy variable α equal to $x_1 - x_2$. In words, the distribution for α is determined by summing over all values of x_1 and x_2 weighted by their joint distribution $p(x_1, x_2)$ but forcing alpha to equal the difference using the dirac delta function. You can use the delta function to do the integral over one of the variables, say x_2 , by solving for x_2 : where the delta function is nonzero $x_2 = x_1 - \alpha$. This leaves us with:

$$p(\alpha) = \int_{\text{all values of } x_1} dx_1 p(x_1, x_1 - \alpha)$$

Note that $p(x_1, x_1 - \alpha)$ will vanish if we go outside the allowed range (which in the pset problem is $(0, \pi)$).

12.1 Application to pset 3 problem 4

In this problem we are given some distribution $p(\nu)$ for a random variable $\nu \geq 0$. We are asked to find the resulting distribution for

$$\lambda \equiv \frac{c}{\nu}. \tag{12.1}$$

(ν is frequency, λ is wavelength, c is the speed of light.)

As in the previous section, we can find this by summing over all values of ν a delta function which sets the right function of ν equal to our new variable:

$$p_\lambda(l) = \int d\nu p_\nu(\nu) \delta\left(l - \frac{c}{\nu}\right). \tag{12.2}$$

Note that I have used subscripts to distinguish $p_\lambda(\lambda)$ and $p_\nu(\nu)$ because *they are different functions*, not just one function evaluated in terms of a different independent variable. This is the real reason for all the business with the dummy variables. The direct relation between these functions is (which follows from (12.2)):

$$|p_\lambda(\lambda)d\lambda| = |p_\nu(\nu)d\nu| \tag{12.3}$$

To see that this follows from (12.2), just change variable of integration in (12.2) from ν to $L = \frac{c}{\nu}$ so $d\nu = -\frac{c}{L^2}dL$ and so

$$p_\lambda(l) = \int_\infty^0 \frac{-cdL}{L^2} p_\nu\left(\frac{c}{L}\right) \delta(l - L) = \int_0^\infty \frac{cdL}{L^2} p_\nu\left(\frac{c}{L}\right) \delta(l - L). \tag{12.4}$$

Now we can easily do the L -integral using the delta function by setting $L = l$ and find

$$p_\lambda(l) = \frac{c}{l^2} p_\nu\left(\frac{c}{l}\right)$$

Comments: (1) given the relation (12.1) the measures are related by

$$d\nu = -c \frac{d\lambda}{\lambda^2}$$

(or $\frac{d\nu}{\nu} = -\frac{d\lambda}{\lambda}$) . The minus sign is because ν increasing means λ decreasing. In the relationship

$$1 = \int_0^\infty d\nu p_\nu(\nu) = \int_0^\infty d\lambda p_\lambda(\lambda)$$

(or in any other relationship between cumulative probabilities) this is taken care of by the fact that the limits of integration are also switched.

(2) One should be careful in using (12.3) in the case that the two variables are not monotonic functions of each other, since this means that there can be more than one value of the original variable which give the same value of the function (*i.e.* the function may not be uniquely invertible). This subtlety is taken care of by either the delta function formula (12.2) or the graphical ABC method.

(3) Notice that the seemingly-pedantic subscripts on the probability distributions are actually very significant:

$$p_\nu(\nu) \neq p_\lambda(c/\nu) \ ;$$

they are *not* the same function! Rather (12.3) is closer to the truth.

(4) The result described here can be considered to be a property of the dirac delta function:

$$\delta(l - f(\nu)) = \delta(f^{-1}(l) - \nu) \frac{1}{|df/d\nu|} \ .$$

Note that the function f does not necessarily have a unique inverse; one must sum over all inverses that exist:

$$\int dl \delta(l - f(\nu)) g(l) = \sum_{\nu_i | f(\nu_i) = l} \frac{1}{|\partial_\nu f(\nu_i)|} g(f(\nu_i))$$

12.2 More examples

In case this is not enough examples for you, there will be one more in the last section of these notes about the Central Limit Theorem. Also, you might try pset 3 problem 3 from the S2008 OCW version of 8.044: <http://ocw.mit.edu/courses/physics/8-044-statistical-physics-i-spring-2008/assignments/ps3.pdf>. Finally, problem 1 of pset 4 is another opportunity.

13 Comments on central limit theorem

[from Kardar chapter 2]

Last time we introduced the characteristic function $\tilde{p}(k)$ of a distribution $p(x)$

$$\tilde{p}(k) = \langle e^{-ikx} \rangle$$

which is its fourier transform. This definition can be generalized to multiple random variables. Given $p(x_1, \dots, x_N)$,

$$\tilde{p}(k_1, \dots, k_N) = \langle e^{-ik_1x_1} e^{-ik_2x_2} \dots e^{-ik_Nx_N} \rangle = \langle e^{-i \sum_{i=1}^N k_i x_i} \rangle$$

which generates correlation functions of the x_i :

$$\langle x_1^{m_1} \dots x_N^{m_N} \rangle = \prod_{i=1}^N (\partial_{-ik_i})^{m_i} \tilde{p}(k_1, \dots, k_N)|_{k_i=0}$$

Suppose we're interested in the sum of these RVs, which KR calls S_N :

$$X \equiv x_1 + \dots + x_N = \sum_{i=1}^N x_i$$

This has a distribution (using the trick described above) given by

$$p(X) = \int dx_1 \dots dx_N p(x_1, \dots, x_N) \delta \left(X - \sum_{i=1}^N x_i \right).$$

Its characteristic function is

$$\tilde{p}_X(k) = \langle e^{-ik \sum_j x_j} \rangle = \int dX e^{-ikX} p(X) = \tilde{p}(k_1 = k_2 \dots = k_N = k).$$

We can use this object to motivate the Central Limit Theorem (CLT).

Consider the log of this object $\ln \tilde{p}_X(k)$; recall that this is the generating function for *cumulants*, which are some convenient combinations of the moments:

$$\ln \tilde{p}_X(k) = -ik \langle X \rangle_c + \frac{(-ik)^2}{2} \langle X^2 \rangle_c + \dots$$

this expression defines the objects $\langle X^m \rangle_c$. By comparing to the taylor expansion of $\ln \tilde{p}$, we find

$$\langle X \rangle_c = \sum_{i=1}^N \langle x_i \rangle_c, \quad \langle X^2 \rangle_c = \sum_{i,j=1}^N \langle x_i x_j \rangle_c,$$

where $\langle x_i x_j \rangle_c$ is defined from the Taylor expansion of $\ln \tilde{p}(k_1 \dots k_N)$.

Consider the case where the RV's are statistically independent and for convenience identically distributed :

$$p(x_1, \dots x_N) = \prod_{i=1}^N p(x_i)$$

(this is sometimes abbreviated 'iid' for 'identically, independently distributed'). Then

$$\tilde{p}(k_1 \dots k_N) = \prod_i \tilde{p}_i(k_i) = \prod_i \tilde{p}(k_i)$$

since the integrals over x_i don't care about each other; in the last step we used the 'identically' part. Further $\tilde{p}_X(k) = \tilde{p}(k)^N$. This means that the 'cross-cumulants' $\langle x_i x_j \rangle_c$ for $i \neq j$ all vanish – the RVs are uncorrelated. So

$$\langle X^m \rangle_c = \sum_{i=1}^N \langle x_i^m \rangle_c = N \langle x^m \rangle_c.$$

Now if we let

$$y \equiv \frac{x - N \langle x \rangle}{\sqrt{N}}$$

then we have $\langle y \rangle = 0$, $\text{Var}(y) = \langle y^2 \rangle_c = \langle x^2 \rangle_c$ but

$$\langle y^m \rangle_c \propto N^{1-m/2}$$

which for $m > 2$ vanishes as $N \rightarrow \infty$. Since the Gaussian distribution is the only one whose only cumulants are the first two, this means that $p(y)$ is Gaussian. This the CLT:

$$p(y) = \frac{1}{\sqrt{2\pi \langle x^2 \rangle_c}} e^{-\frac{y^2}{2 \langle x^2 \rangle_c}}$$

Actually, we didn't need the RVs to be SI in order to obtain this result. We just need that they are not *too* correlated. A sufficient condition is

$$\sum_{i_1 \dots i_m}^N \langle x_{i_1} \dots x_{i_m} \rangle_c \ll \mathcal{O}(N^{m/2})$$

in order for the higher cumulants of y to vanish at large N .

14 Comment about thermodynamic variables which are neither intensive nor extensive

In lecture, extensive thermodynamic variables were defined to be proportional to the number of particles N^1 , while intensive variables go like N^0 . The interesting question arose as to whether there are other alternatives for the dependence on N of thermodynamic variables besides these two powers. If the particles are *free*, *i.e.* do not interact with each other, other things don't arise. More generally, when interactions are important, there are indeed other possibilities. Unfortunately I'm having a hard time coming up with an example that's simple enough to be worth explaining here (the theory of N D0-branes in type IIA superstring theory, which has phases where the energy goes like N^2 , doesn't cut it).

15 1d chain: entropic forces (another application of the binomial distribution and its gaussian peak)

Here is another application of the manipulations Krishna did in lecture on Wednesday Feb. 23, 2011, exhibiting the gaussian nature of the binomial distribution at large N . [From 8.08 pset 1] This problem also has some conceptual overlap with problem 7 of pset 4.

Consider a flexible chain of N links of length a . Each link can point up or down. A ball of mass m is attached to the end of the chain. Every configuration of the chain has the same energy, except for the gravitational energy associated with the height of the ball.

(a) How many configurations of the chain produce a net length L ? Call it $\Omega(L)$.

Let n_\uparrow and n_\downarrow denote the number of links which point up or down.

$$\frac{L}{a} = n_\uparrow - n_\downarrow = 2n_\downarrow - N$$

We used the fact that the total number of links is $N = n_\uparrow + n_\downarrow$. So: $n_\downarrow(L) = \frac{L/a + N}{2}$.

The counting of configurations is the same as in coin flipping or as in the random walk in one dimension.

$$\Omega(L) = \binom{N}{n_\downarrow(L)} = \frac{N!}{\left(\frac{N+L/a}{2}\right)! \left(\frac{N-L/a}{2}\right)!}$$

Note that the length with the maximum number of configurations is $L = 0$. If we ignored the force from the mass, the most probable configuration would be $L = 0$.

(b) Find the energy and entropy of the chain when the length of the chain is L . What happens in the (most likely) limit $L \ll Na$?

$$E = -mgL$$

up to an additive constant which will not matter.

$$S = k_B \ln \Omega.$$

I will work in units where $k_B = 1$ for the rest of this discussion. In the limit $L/a \ll N$ we can do exactly the set of manipulations from the lecture using Stirling :

$$S = \ln \Omega \approx N \ln N - \left(\frac{N}{2} + \frac{L}{2a}\right) \ln \left(\frac{N}{2} + \frac{L}{2a}\right) - \left(\frac{N}{2} - \frac{L}{2a}\right) \ln \left(\frac{N}{2} - \frac{L}{2a}\right)$$

and Taylor expanding the logarithms. To make the comparison with lecture more explicit, let $\epsilon \equiv \frac{L}{2a}$; this is exactly the deviation from the mean $n_\downarrow = \frac{N}{2} + \epsilon = \langle n_\downarrow \rangle + \epsilon$. Again there are many cancellations:

$$\begin{aligned} S &\approx N \ln N - \left(\frac{N}{2} + \epsilon\right) \ln \left(\frac{N}{2} + \epsilon\right) - \left(\frac{N}{2} - \epsilon\right) \ln \left(\frac{N}{2} - \epsilon\right) \\ S &\approx \frac{\epsilon^2}{2N} \left(1 + \mathcal{O}\left(\frac{\epsilon}{N}\right)\right) = -\frac{L^2}{2Na} \left(1 + \mathcal{O}\left(\frac{L}{Na}\right)\right) \end{aligned}$$

I would like to make some remarks. First, the reason that the order- ϵ terms cancel is that we are expanding around the *maximum* of the distribution. The statement that it is the maximum means that the derivative vanishes there; that derivative is exactly the linear term in the Taylor expansion. Second, the nontrivial statement of the Central Limit Theorem here is not just that we can Taylor expand the log of the distribution about the maximum. The nontrivial statement is that the terms of higher order than ϵ^2 become *small* as $N \rightarrow \infty$. It is crucial here that the terms we are neglecting go like $\frac{\epsilon^3}{N^2}$.

(c) Find the temperature T of the chain when the length is L .

$$\frac{1}{T} = \frac{dS}{dE} = \frac{dS/dL}{dE/dL} = \frac{-\frac{2L}{a^2N} + \mathcal{O}\left(\frac{L^2}{a^2N^2}\right)}{-mg} \quad (15.1)$$

(d) Find the length of the chain in terms of T . Show that the length of the chain is proportional to the gravitational force mg for a fixed T and a small force. This means that the chain acts like a harmonic oscillator, *i.e.* satisfies Hooke's law, despite the fact that there is no mechanical restoring force.

For small mg and hence small L , we can just keep the leading term in (22.1)

$$L \approx \frac{a^2 N}{T} mg$$

So Newton's law here is:

$$mg \approx k_{\text{hooke}} L$$

where the spring constant is $k_{\text{hooke}} = Ta^2N$.

16 Comment on limits of integration

Many probability distributions (*e.g.* in pset 4 problem 1) are defined to be zero outside of some specified range. When manipulating these distributions (*e.g.* when convolving them) one must not forget this.

17 Warmup problem

Consider the locus

$$\mathcal{S} \equiv \{(x_1, x_2, x_3) \in \mathbb{R}^3 | 0 = F(x_1, x_2, x_3)\}$$

where the function F is

$$F(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - a^2.$$

0) What is the dimension of this locus?

1) Draw it.

2) For $x_1, x_2, x_3 \in \mathcal{S}$, *i.e.* restricted to the surface, compute

$$\left. \frac{\partial x_1}{\partial x_2} \right|_{x_3} \left. \frac{\partial x_2}{\partial x_3} \right|_{x_1} \left. \frac{\partial x_3}{\partial x_1} \right|_{x_2}$$

The surface can be expressed as

$$x_1 = x_1(x_2, x_3) = \pm \sqrt{a^2 - x_2^2 - x_3^2}$$

with similar expressions interchanging 1, 2, 3. Note that there are three choices of sign, which describe points in the $2^3 = 8$ octants of \mathbb{R}^3 . For simplicity, let's take a point in the octant where $x_{1,2,3} > 0$. Then we have

$$\left. \frac{\partial x_1}{\partial x_2} \right|_{x_3} = \frac{1}{2} \frac{-2x_2}{\sqrt{a^2 - x_2^2 - x_3^2}} = -\frac{x_2}{x_1}$$

$$\frac{\partial x_2}{\partial x_3} \Big|_{x_1} = \frac{1}{2} \frac{-2x_3}{\sqrt{a^2 - x_3^2 - x_1^2}} = -\frac{x_3}{x_2}$$

$$\frac{\partial x_3}{\partial x_1} \Big|_{x_2} = \frac{1}{2} \frac{-2x_1}{\sqrt{a^2 - x_1^2 - x_2^2}} = -\frac{x_1}{x_3}$$

So the product is

$$\frac{\partial x_1}{\partial x_2} \Big|_{x_3} \frac{\partial x_2}{\partial x_3} \Big|_{x_1} \frac{\partial x_3}{\partial x_1} \Big|_{x_2} = (-1)^3 \frac{x_2}{x_1} \frac{x_3}{x_2} \frac{x_1}{x_3} = -1.$$

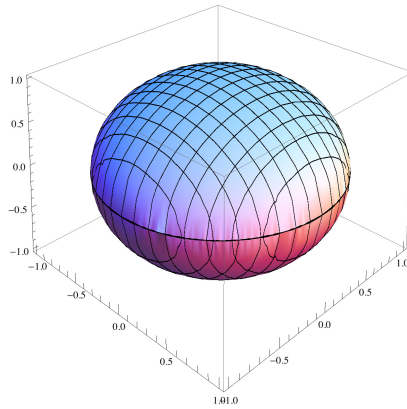


Figure 10: The locus \mathcal{S} for $a = 1$

18 Relations among response functions

[Adkins, page 12 discusses the case of $n = 3$]

Consider n variables satisfying one constraint

$$0 = F(x_1, \dots, x_n), \tag{18.1}$$

so that $n - 1$ of them are independent. This specifies a locus in the n -dimensional space whose dimension is $n - 1$. We can write n different equations imposing the constraint

$$x_1 = x_1(x_2 \dots x_n), x_2 = x_2(x_1, x_3 \dots x_n), \dots, x_n = x_n(x_1 \dots x_{n-1}),$$

each of which encodes the same information about the shape of this $(n - 1)$ -dimensional space. (In the case of $n = 3$, then, (18.1) specifies a surface in three dimensions.) Note that not all of these functions will be single-valued – a simple example is the case $x_1^2 = x_2$, in which there are two values of $x_1 = \pm\sqrt{x_2}$ for each value of x_2 (for $x_2 > 0$,

or no real values if $x_2 < 0$). The relations we study below are true *locally*, *i.e.* pick a branch of solutions and stick with it.

The variation of any given variable as we move along the constraint surface can be expressed in terms of the variations of the others:

$$dx_1 = \sum_{j \neq 1} \frac{\partial x_1}{\partial x_j} \Big|_{x_{l \neq 1, j}} dx^j \quad (18.2)$$

$$dx_2 = \sum_{j \neq 2} \frac{\partial x_2}{\partial x_j} \Big|_{x_{l \neq 1, j}} dx^j \quad (18.3)$$

Note that you can actually ignore the annoying $|_{x_{l \neq 1, j}}$ bits here– they always express that all the other variables are fixed besides the two that are involved in the derivative. So I’m going to suppress them here to make the equations look better – you have to remember that they are there. Now substitute in (18.2) using (18.3) for dx_2 :

$$dx_1 = \sum_{j \neq 1, 2} \frac{\partial x_1}{\partial x_j} dx^j + \frac{\partial x_1}{\partial x_2} \left(\sum_{j \neq 1, 2} \frac{\partial x_2}{\partial x_j} dx^j + \frac{\partial x_2}{\partial x_1} dx_1 \right)$$

Now we group terms together:

$$0 = dx_1 \left(-1 + \frac{\partial x_1}{\partial x_2} \frac{\partial x_2}{\partial x_1} \right) + \sum_{j \neq 1, 2} \left(\frac{\partial x_1}{\partial x_j} + \frac{\partial x_1}{\partial x_2} \frac{\partial x_2}{\partial x_j} \right)$$

Now since we can vary x_1 and $x_3, x_4 \dots x_n$ independently, this is actually $n - 1$ equations. Varying only x_1 we learn that:

$$-1 + \frac{\partial x_1}{\partial x_2} \frac{\partial x_2}{\partial x_1}$$

i.e.

$$\frac{\partial x_1}{\partial x_2} \Big|_{\text{all others fixed}} = \frac{1}{\frac{\partial x_2}{\partial x_1} \Big|_{\text{all others fixed}}}$$

Since 1, 2 aren’t special, we have

$$\frac{\partial x_i}{\partial x_j} \Big|_{\text{all others fixed}} = \frac{1}{\frac{\partial x_j}{\partial x_i} \Big|_{\text{all others fixed}}} \quad (18.4)$$

for any distinct i, j . This is sometimes called the “reciprocal theorem”.

Varying any of the x_j with $j = 3..n$, we learn that:

$$\forall j \neq 1, 2, \quad -1 = \frac{\partial x_1}{\partial x_2} \frac{\partial x_2}{\partial x_j} \frac{\partial x_j}{\partial x_1}$$

Realizing that there isn't anything special about x_1, x_2 , this says:

$$-1 = \frac{\partial x_i}{\partial x_j} \frac{\partial x_j}{\partial x_k} \frac{\partial x_k}{\partial x_i} \quad (18.5)$$

for any i, j, k distinct. Note that the RHS is dimensionless because each of $x_{i,j,k}$ appears once in the top and once in the bottom. This is sometimes called the “reciprocity theorem”.

18.1 Comparison to 1d chain rule

Let's go back to the case of three variables for simplicity, and call them $(x_1, x_2, x_3) = (x, y, z)$. If we combine this last relation (18.5) with the reciprocal relation (18.4), we have

$$\frac{\partial y}{\partial x} = -\frac{\partial y}{\partial z} \frac{\partial z}{\partial x}.$$

This way of writing (18.5) makes the minus sign seem to conflict with the chain rule in single-variable calculus, $\frac{dY}{dX} = \frac{dY}{dZ} \frac{dZ}{dX}$. There is actually no conflict, because the latter formula applies to a different situation, namely where X, Y, Z each determine each other, *i.e.* we have *two* independent relations $X = X(Y)$ AND $Y = Y(Z)$ among the three variables (which specifies a *curve* in space, rather than a surface).

But this makes the sign hard to remember.

18.2 An even simpler example

To check that this funny-seeming sign is really there, let's do a simple example. Take

$$0 = F(x, y, z) = x + y + z.$$

Then

$$x(y, z) = -y - z, \quad y(x, z) = -x - z, \quad z(x, y) = -x - y$$

$$\frac{\partial x}{\partial y} = -1, \quad \frac{\partial y}{\partial z} = -1, \quad \frac{\partial z}{\partial x} = -1$$

So indeed their product is

$$\frac{\partial x}{\partial y} \frac{\partial y}{\partial z} \frac{\partial z}{\partial x} = -1.$$

19 Exact differentials and second derivatives

I don't have much to add here beyond Adkin's discussion on pages 14-15, except the following.

1) Beware that if $x = x(y, z)$ is a multi-valued function, trouble can arise in Adkins' argument. If the two paths end up on different branches of the function, the results need not agree. The mixed partials only disagree at the branch point.

2) The condition that

$$F_x dx + F_y dy$$

is an exact differential, namely that

$$\partial_y F_x - \partial_x F_y = 0$$

is a condition familiar from other contexts (e.g. electricity and magnetism) as the condition that the 2d vector field $\vec{F} \equiv F_x \hat{x} + F_y \hat{y}$ be curl-free,

$$\vec{\nabla} \times \vec{F} = 0. \tag{19.1}$$

That is in turn equivalent (at least in flat space with no holes) to the condition that it be someone's gradient

$$\vec{F} = \vec{\nabla} f = \hat{x} \partial_x f + \hat{y} \partial_y f.$$

This is the statement that

$$F_x dx + F_y dy = df.$$

The condition (19.1) also says that line integrals of \vec{F} are path-independent:

$$\int_1^2 (F_x dx + F_y dy) = f(2) - f(1),$$

and integrals over closed loops vanish

$$\oint (F_x dx + F_y dy) = 0.$$

20 1d chain: entropic forces (another application of the binomial distribution and its gaussian peak), part 1

Here is another application of the manipulations Krishna did in lecture on Wednesday Feb. 23, 2011, exhibiting the gaussian nature of the binomial distribution at large N .

[From 8.08 pset 1] This problem also has some conceptual overlap with problem 7 of pset 4.

Consider a flexible chain of N links of length a . Each link can point up or down. Later on, we will attach a ball of mass m to the end of the chain. Every configuration of the chain has the same energy, except for the gravitational energy associated with the height of the ball.

(a) How many configurations of the chain produce a net length L ? Call it $\Omega(L)$.

Let n_{\uparrow} and n_{\downarrow} denote the number of links which point up or down.

$$\frac{L}{a} = n_{\uparrow} - n_{\downarrow} = 2n_{\downarrow} - N$$

We used the fact that the total number of links is $N = n_{\uparrow} + n_{\downarrow}$. So: $n_{\downarrow}(L) = \frac{L/a + N}{2}$.

The counting of configurations is the same as in coin flipping or as in the random walk in one dimension.

$$\Omega(L) = \binom{N}{n_{\downarrow}(L)} = \frac{N!}{\left(\frac{N+L/a}{2}\right)! \left(\frac{N-L/a}{2}\right)!}$$

Note that the length with the maximum number of configurations is $L = 0$. If we ignored the force from the mass, the most probable configuration would be $L = 0$.

(b) What is the probability distribution for L ? Use the Central Limit Theorem.

The information about the probability distribution for L is contained in the previous calculation of the number of configurations, since

$$p(L) = \frac{\Omega(L)}{2^N}$$

i.e. $p(L)$ is the number of configurations where the length of the chain is L divided by the total number of configurations of a chain of length N , namely 2^N .

Having already done this the hard way by actually counting configurations, let's redo it the easy way using the Central Limit Theorem. This theorem implies that the sum of N SI RVs $X \equiv \sum_{i=1}^N x_i$ is governed by a Gaussian probability distribution. This fact is extremely useful because a Gaussian distribution is determined by its mean and its variance, and we know those for X , namely

$$\langle X \rangle = N \langle x \rangle, \quad \text{Var}(X) = N \text{Var}(x).$$

⁵ That is, we can figure out the mean and variance for X in terms of the mean and variance in the case of $N = 1$ where there is just one variable, which is in general easy.

⁵To recap,

$$\langle X \rangle = \langle x_1 + \dots + x_N \rangle = \langle x_1 \rangle + \dots + \langle x_N \rangle$$

If there were only one link of the chain, $N = 1$, the mean value of L would be zero and the variance would be

$$\text{Var}_{N=1}(L) = \langle L^2 \rangle_{N=1} - \langle L \rangle_{N=1}^2 = \frac{1}{2}a^2 + \frac{1}{2}a^2 = a^2.$$

So the CLT says

$$p(L) = \frac{1}{\sqrt{2\pi N \text{Var}(x)}} e^{-\frac{(L-N\langle x \rangle)^2}{2N \text{Var}(x)}} = \frac{1}{\sqrt{2\pi N a^2}} e^{-\frac{L^2}{2N a^2}}$$

Note that this equals sign really means “can be well represented by” since the values of L that are actually achieved are integer multiples of a ; for large N , this spacing becomes negligible compared to the width of the envelope $\sim \sqrt{N}a$.

Next time we will think about the consequences of this if we hang a mass on the chain and it is in thermal equilibrium.

no matter whether the x_i are SI, just because taking expectations is a linear operation. If the x_i are also identically distributed, this is $\langle X \rangle = N \langle x \rangle$.

$$\text{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2.$$

$$\langle X^2 \rangle = \left\langle \sum_{i=1}^N x_i^2 + 2 \sum_{i<j} \langle x_i x_j \rangle \right\rangle.$$

There are N terms of the first kind, and $\frac{1}{2}N(N-1)$ terms of the second kind, each of which is of the form $\langle x_1 x_2 \rangle = \langle x_1 \rangle \langle x_2 \rangle$ by the SI assumption. So, assuming the x_i are iid,

$$\langle X^2 \rangle = n \langle x^2 \rangle + n(n-1) \langle x \rangle^2$$

and

$$\text{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = n \langle x^2 \rangle + n(n-1) \langle x \rangle^2 - (n \langle x \rangle)^2 = n \text{Var}(X).$$

Here are two quick applications of these results:

1) Random walk in one dimension: The average endpoint after N steps is $\langle X \rangle = N \langle x \rangle = 0$, but the spread of the distribution of the endpoint after N steps is

$$\text{Var}(X) = N \langle x^2 \rangle = NL^2$$

where L is the step length. This says that the RMS displacement from the origin is $\sqrt{\text{Var}(X)} = \sqrt{NL}$; if we imagine each step takes a time t_0 , the RMS displacement after time t satisfies

$$\Delta x(t) = \sqrt{\frac{t}{t_0}} L$$

which is characteristic of diffusion.

2) In doing an experiment, why is it a good idea to take many measurements and average the result? We have $\text{Var}(\sum_i x_i) = N \text{Var}(x)$. The average of the results is $\bar{x} \equiv \frac{1}{N} \sum_{i=1}^N x_i$; since the variance satisfies $\text{Var}(ay + b) = a^2 \text{Var}(y)$, we have $\text{Var}(\bar{x}) = \frac{1}{N} \text{Var}(x)$, so the statistical error in the average shrinks (compared to the statistical error in one measurement) as N grows. Note that this says nothing about systematic errors.

21 Warmup problem: exact and inexact differentials

[from Greiner, *Thermodynamics and Statistical Mechanics*]

Suppose someone gives you a differential

$$F = \vec{F} \cdot d\vec{x} = F_x(x, y)dx + F_y(x, y)dy$$

which is not exact ($F \neq df$), such as

$$\vec{F} \cdot d\vec{x} = yxdx + x^2dy, \quad (21.1)$$

as you can see by computing

$$\partial_y F_x \neq \partial_x F_y.$$

Sometimes it is possible to find another differential proportional to the original one which *is* exact, *i.e.* there exists $g(x, y)$ such that

$$gF = df. \quad (21.2)$$

In fact there can be many possible g , which is naturally called an “integrating factor”. Let’s find one such g for the example (21.1).

The desideratum (21.2) implies that gF is closed, *i.e.*

$$\partial_x(gF_y) = \partial_y(gF_x).$$

Given F , this is a PDE for g . For our special case, it is

$$\partial_x(gx^2) = \partial_y(gxy)$$

which says

$$g + x\partial_x g = y\partial_y g.$$

It can be solved by separation of variables: $g = g_1(x)g_2(y)$, which gives

$$g_1 g_2 + x g_1' g_2 = y g_1 g_2'$$
$$1 + x \frac{\partial_x g_1(x)}{g_1(x)} = y \frac{\partial_y g_2(y)}{g_2(y)} = C.$$

Since the LHS depends only on x and the RHS depends only on y , the BHS is a constant, C . The solution of

$$y \frac{\partial_y g_2(y)}{g_2(y)} = C$$

is $\ln g_2 = C \ln y + C_2$, for a constant C_2 or

$$g_2 = e^{C_2} y^C.$$

Similarly

$$1 + x \frac{\partial_x g_1(x)}{g_1(x)} = C$$

is solved by $g_1 = e^{C_1} x^{C-1}$. But we only care about the product $g_1 g_2$, and we just need one solution of this equation, so let's set $C_1 = C_2 = C = 0$, and we have

$$g(x, y) = x^{C-1} y^C = \frac{1}{x}.$$

Indeed

$$\frac{1}{x} F = y dx + x dy = d(xy + a)$$

(where a is some constant) is exact. And so is

$$\frac{(xy)^C}{x} F = (xy)^C (y dx + x dy) = d \frac{(xy)^{C+1}}{C+1}$$

for any C .

I mention this point because we've seen that work is *not* an exact differential, that is dW depends on what process we do. On the other hand, for a 'hydrostatic system', *i.e.* a fluid in a container, we have

$$dW = P dV$$

where dV is certainly an exact differential – it is the change in the volume of the system and the volume is a state variable. So here

$$\frac{dW}{P} = dV$$

is exact.

Similarly, heat dQ is not exact. Soon we will see that it can also be written as a state function times an exact differential:

$$dQ = T dS.$$

A challenge: try to come up with an example of a differential for which this procedure cannot be done. You may need more independent variables.

22 1d chain: entropic forces (another application of the binomial distribution and its gaussian peak)

Here is another application of the manipulations Krishna did in lecture on Wednesday Feb. 23, 2011, exhibiting the gaussian nature of the binomial distribution at large N . [From 8.08 pset 1; it also makes an appearance on 8.044 pset 5b. This problem also has some conceptual overlap with problem 7 of pset 4.] Note that I am giving here a taste of things to come and violating a bit the injunction against talking about macro and micro perspectives at the same time. Please be aware of this and forgive me.

Consider a flexible chain of N links of length a . Each link can point up or down. A ball of mass m is attached to the end of the chain. Every configuration of the chain has the same energy, except for the gravitational energy associated with the height of the ball.

(a) How many configurations of the chain produce a net length L ? Call it $\Omega(L)$. [Recall that on Tuesday, March 1, 2011, we discussed $p(L)$, where p and Ω are related by

$$p(L) = \frac{\Omega(L)}{2^N} = \frac{e^{S(L)}}{2^N},$$

as an application of the CLT. Note that we needn't worry too much about the (N -dependent) constant in front because it is specified by normalizing the distribution $\sum_L p(L) = 1$ anyway.]

Let n_\uparrow and n_\downarrow denote the number of links which point up or down.

$$\frac{L}{a} = n_\uparrow - n_\downarrow = 2n_\downarrow - N$$

We used the fact that the total number of links is $N = n_\uparrow + n_\downarrow$. So: $n_\downarrow(L) = \frac{L/a + N}{2}$.

The counting of configurations is the same as in coin flipping or as in the random walk in one dimension.

$$\Omega(L) = \binom{N}{n_\downarrow(L)} = \frac{N!}{\left(\frac{N+L/a}{2}\right)! \left(\frac{N-L/a}{2}\right)!}$$

Note that the length with the maximum number of configurations is $L = 0$. If we ignored the force from the mass, the most probable configuration would be $L = 0$.

(b) Find the entropy of the chain when the length of the chain is L . What happens in the (most likely) limit $L \ll Na$?

$$S = k_B \ln \Omega.$$

I will work in units where $k_B = 1$ for the rest of this discussion. In the limit $L/a \ll N$ we can do exactly the set of manipulations from the lecture using Stirling :

$$S = \ln \Omega \approx N \ln N - \left(\frac{N}{2} + \frac{L}{2a}\right) \ln \left(\frac{N}{2} + \frac{L}{2a}\right) - \left(\frac{N}{2} - \frac{L}{2a}\right) \ln \left(\frac{N}{2} - \frac{L}{2a}\right)$$

and Taylor expanding the logarithms. To make the comparison with lecture more explicit, let $\epsilon \equiv \frac{L}{2a}$; this is exactly the deviation from the mean $n_\downarrow = \frac{N}{2} + \epsilon = \langle n_\downarrow \rangle + \epsilon$. Again there are many cancellations:

$$\begin{aligned} S &\approx N \ln N - \left(\frac{N}{2} + \epsilon\right) \ln \left(\frac{N}{2} + \epsilon\right) - \left(\frac{N}{2} - \epsilon\right) \ln \left(\frac{N}{2} - \epsilon\right) \\ S &\approx -\frac{2\epsilon^2}{N} \left(1 + \mathcal{O}\left(\frac{\epsilon}{N}\right)\right) = -\frac{L^2}{2Na} \left(1 + \mathcal{O}\left(\frac{L}{Na}\right)\right) \end{aligned}$$

I would like to make some remarks. First, the reason that the order- ϵ terms cancel is that we are expanding around the *maximum* of the distribution. The statement that it is the maximum means that the derivative vanishes there; that derivative is exactly the linear term in the Taylor expansion. Second, the nontrivial statement of the Central Limit Theorem here is not just that we can Taylor expand the log of the distribution about the maximum. The nontrivial statement is that the terms of higher order than ϵ^2 become small as $N \rightarrow \infty$. It is crucial here that the terms we are neglecting go like $\frac{\epsilon^3}{N^2}$.

(c) Find the energy of the chain when the length of the chain is L .

$$E = -mgL$$

up to an additive constant which will not matter.

(d) Assuming thermal equilibrium, find the temperature T of the chain when the length is L .

Here we will cheat a little and use a formula that we saw briefly in the preview chapter of this course:

$$\frac{1}{T} = \frac{dS}{dE} = \frac{dS/dL}{dE/dL} = \frac{-\frac{L}{a^2 N} + \mathcal{O}\left(\frac{L^2}{a^2 N^2}\right)}{-mg} \quad (22.1)$$

(d) Find the length of the chain in terms of T . Show that the length of the chain is proportional to the gravitational force mg for a fixed T and a small force. This means that the chain acts like a harmonic oscillator, *i.e.* satisfies Hooke's law, despite the fact that there is no mechanical restoring force.

For small mg and hence small L , we can just keep the leading term in (22.1)

$$L \approx \frac{a^2 N}{2T} mg$$

So Newton's law here is:

$$mg \approx k_{\text{hooke}} L$$

where the spring constant is $k_{\text{hooke}} = \frac{2T}{a^2 N}$.

Note that the fact that the force *about equilibrium* is linear in the extent of the chain is not surprising – this is just the statement that it has a Taylor expansion; the real consequence of having a large chain of length N (making the CLT relevant) is that the ‘anharmonic terms’ (*i.e.* the terms in the force that go like L^2 and higher powers) are suppressed by powers of N .

To recap, I had several purposes in discussing this example:

- It's an opportunity to apply the CLT, along the same lines as in pset 4.
- It describes a physical situation similar to the one discussed from a thermodynamic viewpoint in the last problem of pset 4.
- The discussion of microscopic counting of configurations of the chain was an opportunity to emphasize the significant steps in the derivation of the statement that the binomial distribution becomes gaussian at large N .
- This is an example of an ‘entropic force’: the restoring force on the chain arises exclusively from the fact that the entropy is larger when it is not stretched.

23 Adiabats and entropy: a review problem

Given the equation of state of a hydrostatic system $P = P(V, T)$, and the internal energy $U = U(T, V)$, let's find the entropy of the system $S(T, V)$. Two examples which are relevant to our present purposes (recent lectures, pset 5a, review for the exam on Friday) where we can be more specific are

a) the ideal gas where

$$\text{ideal gas: } PV = NkT, \quad U(V, T) = \frac{3}{2}NkT.$$

b) the VdW gas where

$$\text{VdW gas: } \left(P + \frac{a}{V^2}\right)(V - b) = NkT, \quad U(V, T) = \frac{3}{2}NkT - \frac{a}{V}.$$

In general, the first law says

$$dU = \delta Q + \delta W .$$

Applied to a hydrostatic system, it is

$$dU = \delta Q - PdV .$$

In general, the heat absorbed by our system may be written as

$$\delta Q = TdS$$

where S is a state variable called the entropy. I will motivate the claim that the quantity S so defined is a state variable below in section 23.4.

So we have

$$TdS = +PdV + dU = PdV + (\partial_T U|_V dT + \partial_V U|_T dV) = (P + \partial_V U|_T) dV + \partial_T U|_V dT = (P + \partial_V U|_T) dV + C_V dT$$

Dividing this by T , we have an expression for the differential dS :

$$dS = \frac{1}{T} (P + \partial_V U|_T) dV + \frac{C_V}{T} dT \quad (23.1)$$

(which I claim is exact on general grounds; this claim implies some relation between $\partial_V C_V$ and T -derivatives of the thing multiplying dV).

To make further progress, let's think about the examples.

23.1 Entropy of ideal gas

For the ideal gas, (23.1) becomes:

$$dS = \frac{1}{T} \left(\frac{NkT}{V} + 0 \right) dV + \frac{3}{2} Nk dT = Nk \frac{dV}{V} + \frac{3}{2} Nk \frac{dT}{T} .$$

First let's check that it's exact:

$$\partial_T (\partial_V S)_T = 0 = \partial_V (\partial_T S)_V .$$

I like to integrate such things by picking path of line segments along the coordinate axes, starting from some arbitrary point. Let's move from (T_0, V_0) to (T, V_0) and then to (T, V) . This gives

$$S(T, V) = S(T_0, V_0) + \int_{T_0}^T dT' (\partial_T S(T', V_0))_V + \int_{V_0}^V dV' (\partial_V S(T, V'))_T$$

$$= S(T_0, V_0) + \int_{T_0}^T dT' \frac{3Nk}{2T'} + \int_{V_0}^V dV' \frac{Nk}{V'} = S(T_0, V_0) + Nk \left(\frac{3}{2} \ln \frac{T}{T_0} + \ln \frac{V}{V_0} \right) = \text{const} + Nk \ln T^{3/2} V.$$

We can and should check that indeed

$$S(T, V) = \text{const} + Nk \ln T^{3/2} V \quad (23.2)$$

reproduces the expression for dS above. I derive some enjoyment from the fact that (23.2) is sometimes called ‘the Sackur-Tetrode equation’.

Notice that along an adiabat, $dQ = TdS = 0$, entropy is constant. This is consistent with our previous information that $PV^\gamma = \text{const}$, since here $\gamma = \frac{C_P}{C_V} = 5/3$ so

$$T^{3/2} V = \text{const} \Leftrightarrow (PV)^3 / 2V = \text{const} \Leftrightarrow P^3 V^5 = \text{const} = \text{const} \Leftrightarrow PV^{5/3} = \text{const}.$$

In the second step we used the ideal gas equation of state.

23.2 Entropy of VdW gas

Now we have

$$\text{VdW gas:} \quad \left(P + \frac{a}{V^2} \right) (V - b) = NkT, \quad U(V, T) = \frac{3}{2} NkT - \frac{a}{V}.$$

The constant b reduces the volume available to the particles because of the fact that they repel each other at very small separations. So (23.1) becomes:

$$dS = \frac{1}{T} \left[\left(\frac{NkT}{V-b} - \frac{a}{V^2} + \frac{a}{V^2} \right) dV + cNk dT \right] = \left(\frac{Nk}{V-b} \right) dV + \frac{cNk}{T} dT.$$

First check equality of mixed partials:

$$\partial_T (\partial_V S)_T = 0 = \partial_V (\partial_T S)_V \quad ;$$

both are still zero.

Following the same path as before,

$$\begin{aligned} S(T, V) &= S(T_0, V_0) + \int_{T_0}^T dT' (\partial_T S(T', V_0))_V + \int_{V_0}^V dV' (\partial_V S(T, V'))_T \\ &= S(T_0, V_0) + \int_{T_0}^T dT' \frac{cNk}{T'} + \int_{V_0}^V dV' \frac{Nk}{V' - b} = \text{const} + Nk \ln T^c (V - b) \end{aligned}$$

This differs from the ideal expression above in two ways. First, the exponent $T^{3/2}$ became T^c ; this is not because the gas is not ideal, but just because I called the constant specific heat here $C_V = \partial_T U|_V = cNk$ instead of $\frac{3}{2}Nk$ as above; a non-monatomic ideal gas can indeed have other values of c . The second difference is the appearance of b ; this means that it is

$$T^c (V - b) \propto e^S$$

which is constant along an adiabat for a VdW gas.

23.3 Carnot cycle in T, S plane

[problem 3 from 8.044 S2009 First Test]

Recall that the Carnot cycle takes a substance through a sequence of four quasistatic processes, forming a closed loop, and alternating between isotherms and adiabats. Draw the Carnot cycle in the TS (temperature-entropy) plane.

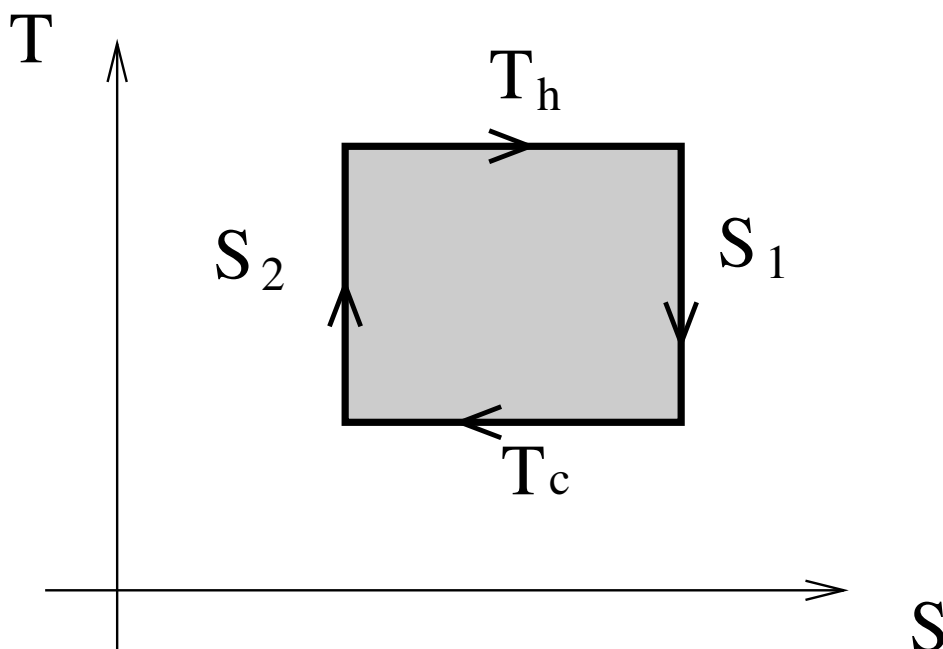


Figure 11: A Carnot cycle in the temperature-entropy plane.

The heat put *into* the engine during the whole cycle is

$$\Delta Q_{\text{total}} = \oint T dS ;$$

the adiabats do not contribute to this integral (even for non-ideal gases).

$$\Delta Q_{\text{total}} = \oint T dS = \Delta Q_{T_h} + \Delta Q_{T_c}$$

This integral is extremely easy to do for the Carnot cycle – it is the area of the rectangle in figure ??:

$$\Delta Q_{\text{total}} = T_h(S_1 - S_2) + T_c(S_2 - S_1) = (T_h - T_c)(S_1 - S_2).$$

The sign is correct when the cycle is traversed in the indicated direction. In that case (which is when the engine *does* work on its surroundings, $\Delta W_{\text{total}} < 0$), ΔQ_{total} is positive because $T_h > T_c$, $S_1 > S_2$.

For the ideal gas $S = a \ln T^{3/2} V$ for a positive constant a ; this means that the adiabat with larger V has higher entropy.

The fact that the process makes a cycle means that

$$0 = \Delta U = \Delta W_{\text{total}} + \Delta Q_{\text{total}} = - \oint P dV + \oint T dS$$

Note that for an ideal gas, the contribution from the isotherms to $\oint(-PdV + TdS)$ vanishes, since the internal energy is constant when T is constant.

23.4 Motivating entropy as a state variable

[Reichl, page 26]

So far we've just assumed the relation $dQ = TdS$, and claimed that S is a state variable. Krishna is going to derive this from statistical mechanics. The following is one way to see just from thermodynamic considerations that entropy is a state variable, *i.e.* that $dS = \frac{dQ}{T}$ is an exact differential.

In a Carnot cycle,

$$0 = \Delta U_{\text{total}} = \Delta Q_{\text{total}} + \Delta W_{\text{total}}.$$

The efficiency is

$$\eta \equiv \frac{\Delta W_{\text{total}}}{\Delta Q_{\text{hot}}} = \frac{\Delta Q_{\text{hot}} + \Delta Q_{\text{cold}}}{\Delta Q_{\text{hot}}} = 1 + \frac{\Delta Q_{\text{cold}}}{\Delta Q_{\text{hot}}}$$

On the other hand, we have shown that $\eta = 1 - \frac{T_c}{T_h}$, so we have

$$0 = \frac{\Delta Q_{\text{hot}}}{T_h} + \frac{\Delta Q_{\text{cold}}}{T_c} = \oint_{\text{carnot}} \frac{dQ}{T}.$$

In the last step we used again the fact that no heat is exchanged during the adiabatic steps.

On the other hand, one can make an arbitrary reversible cycle out of lots of little Carnot cycles. This means that

$$0 = \oint_{\text{any reversible quasistatic cycle}} \frac{dQ}{T}$$

in general, which means $\frac{dQ}{T}$ is exact.

24 Cooling a fluid: some exercises using the first law

[big Reif (*Fundamentals of Statistical and Thermal Physics*), §5.9, 5.10; Greiner, chapter 4]

Last time we talked briefly about adiabatic free expansion of a gas. We noticed that an *ideal* gas does not change its temperature in this process.

24.1 Joule-Thomson Throttling

In practice, a better way to cool fluids (which is used to liquefy gases) is the following procedure introduced by Joule and Thomson (= Lord Kelvin). This procedure also has the advantage of avoiding the intermediate unspeakable non-equilibrium situation.

The idea is to set up a steady flow of gas down an insulating pipe. In the path of the gas is a porous plug, which maintains a pressure differential. On the left, upstream side we have P_1, T_1 , on the right we have P_2 and we'd like to know T_2 .

Consider a fixed mass of gas between two points A, B along the pipe. As the chunk moves, it does work on the gas downstream of it, and it gets work done on it by the gas upstream of it. Think about the work done on this chunk of gas between the time right as it is about to enter the plug and the time right after it has all left the plug. Initially it has volume V_1 , and the gas upstream of it is doing work at pressure P_1 to decrease this volume to zero; this is an amount of work $P_1 V_1$ on the gas. The chunk itself does work against a pressure P_2 on the right to produce a final volume of V_2 , which is work $-P_2 V_2$ done ON the gas. The net work done on this chunk is

$$W = -P_2 V_2 + P_1 V_1.$$

On the other hand, no heat is absorbed because we have assumed steady state and insulating walls. So the first law says

$$0 = Q_{\text{into gas}} = \Delta U_{\text{of gas}} - W_{\text{on gas}}.$$

Note the signs (and note that Reif uses the opposite convention, using W to mean the work done *by* the gas). The change in the internal energy is

$$\Delta U = U_2 - U_1 = U(T_2, P_2) - U(T_1, P_1).$$

So

$$\Delta U = \Delta(PV)$$

or

$$U_1 + P_1V_1 = U_2 + P_2V_2 \quad i.e. H_1 = H_2$$

where $H \equiv U + PV$ is the enthalpy.

Again for an ideal gas this is boring since

$$H = U + PV = U(T) + NkT$$

depends only on T so $H(T_2) = H(T_1)$ just says the temperature is constant and nothing happens. So again we need some interactions in our gas to get an effect. The effect we want is a change in T as we change the pressure, *i.e.* a nonzero value of the “Joule-Thomson coefficient”

$$\mu \equiv \left. \frac{\partial T}{\partial P} \right|_H .$$

$\mu > 0$ ($\mu < 0$) means the gas cools (warms) as we decrease the pressure $P_2 < P_1$.

Let’s find a useful expression for μ . Recall that

$$dU = \delta Q - PdV \quad \longrightarrow \quad dH = d(U + PV) = \delta Q + VdP.$$

Thinking of $H = H(T, P)$ as function of the independent variables T, P , we have

$$dH = \left. \frac{\partial H}{\partial T} \right|_P dT + \left. \frac{\partial H}{\partial P} \right|_T dP.$$

At constant enthalpy, we set this to zero and learn

$$\mu = \left. \frac{\partial T}{\partial P} \right|_H = - \frac{\left. \frac{\partial H}{\partial P} \right|_T}{\left. \frac{\partial H}{\partial T} \right|_P}.$$

We’ve seen in lecture that the thing in the denominator is

$$C_P = \left. \frac{\partial H}{\partial T} \right|_P$$

the specific heat at constant pressure.

We’ll have to work on the numerator some more to relate it to familiar quantities. This will be much easier a little later in the course, so we will come back to it.

25 Comments on pset 5b, problem 2: Rückhardt’s experiment

In recitation, we discussed this problem using Newton’s Law.

Part d) says that we neglected the temperature variation in the earlier parts of the problem. We may not have discussed the temperature variation explicitly, but indeed the temperature must vary. To see this, note that when the ball is at the lowest point of its oscillation, the energy is stored in the internal energy of the gas. Because we are discussing an ideal gas, the internal energy only depends on T , not directly on the volume V . The temperature, however, depends on z : when we adiabatically compress the gas, its internal energy increases via an increase in temperature – the gas molecules move faster. In part d) you will show that this crucial temperature variation is nevertheless numerically not very significant.

The problem can actually be solved using conservation of energy. The total potential energy, as a function of the height of the ball (this does not include kinetic energy of the ball) is:

$$U(z) = -mgz + U_{\text{gas}}(T(z)) = -mgz + cNkT(z).$$

Here c is a constant, which is $3/2$ for a monatomic ideal gas. Since the specific heat is $C_V = \partial_T U_{\text{gas}}|_V = cNk$, and $C_P = C_V + 1$, we have

$$\gamma = \frac{C_P}{C_V} = \frac{c+1}{c} = \frac{5}{3}.$$

The temperature can be inferred from the adiabatic relation $PV^\gamma = \text{const}$ combined with the ideal gas law, $PV = NkT$:

$$P_{\text{eq}} V_{\text{eq}}^\gamma = P(z) V(z)^\gamma = NkT(z) V(z)^{\gamma-1}.$$

here $P_{\text{eq}}, V_{\text{eq}}$ are the values of the temperature and pressure when the ball is in equilibrium. Let the equilibrium position of the ball be $z_{\text{eq}} = 0$; then $V(z) = V_{\text{eq}} + Az$. Therefore:

$$NkT(z) = P_{\text{eq}} V_{\text{eq}}^\gamma V(z)^{1-\gamma}.$$

Therefore

$$U(z) = mgz + cP_{\text{eq}} V_{\text{eq}}^\gamma (V_{\text{eq}} + Az)^{1-\gamma}.$$

Equilibrium requires that the net force on the ball vanish; this means

$$0 = \partial_z U(z_{\text{eq}}) = mg + (1 - \gamma)cP_{\text{eq}} V_{\text{eq}}^\gamma V_{\text{eq}}^{-\gamma} A = mg - P_{\text{eq}} A$$

where we used $1 - \gamma = 1 - \frac{5}{3} = -\frac{2}{3}$. This agrees with the formula for the equilibrium pressure given on the statement of the problem set. Away from equilibrium,

$$m\ddot{z} = -U'(z) = -\left(U'(z_{\text{eq}}) + \frac{1}{2}(z - z_{\text{eq}})U''(z_{\text{eq}}) + \dots \right)$$

which when linearized about equilibrium gives a harmonic oscillator which you can show has the correct frequency.

26 S2008 first test, last problem.

Some of you may have tried the first test from 8.044 S2008, which is available on OCW. The third problem on this exam gives us the heat capacity C_V and a relation between P, V, T for a non-ideal gas, and asks us to find an expression for the internal energy.

I don't know how to do this problem without using a formula from the formula sheet of that exam (or equivalent information which we have not yet covered in 8.044).

For those of you who tried this problem, a derivation of the offending formula is the following. It uses a Maxwell relation, which is a relationship between response functions valid in equilibrium which we'll see in a few weeks.

On the one hand, calculus says:

$$dU = C_V dT + \partial_V U|_T dV.$$

On the other hand, the first law is

$$dU = T dS - P dV.$$

Now expand the differential dS as

$$dS = \partial_T|_V dT + \partial_V S|_T dV$$

to turn the first law into

$$dU = T \partial_T S|_V dT + (T \partial_V|_T S - P) dV.$$

Equating the two resulting expressions for dU we learn: $C_V = T \partial_T S|_V$ and

$$\partial_V U|_T = T \partial_V|_T S - P.$$

Finally, a Maxwell relation says:

$$\partial_V|_T S = \partial_T|_V P.$$

Therefore:

$$\partial_V U|_T = T \partial_T|_V P - P.$$

27 A microcanonical example: Classical magnetic moments

[8.044 S2003 pset 6 number 1]

This problem and its solution are available [here on OCW](#).

Note that in this problem, we ignored the kinetic energy of the moments, which would give an extra contribution to the energy

$$\Delta E = \frac{L^2}{2I}$$

where L is the angular momentum and I is the moment of inertia. This is appropriate in a limit where I is large.

27.1 Whence the expression for $\Omega(M, N)$?

I will add one more part to the problem:

h) Where did the expression for $\Omega(M, N)$ come from?

Notice that the probability of finding a particular magnetization is

$$P(M, N) = \frac{\Omega(M, N)}{\Omega(N)} \propto e^{-\frac{(M-\langle M \rangle)^2}{2\sigma_M}}$$

where $\Omega(N) \equiv \sum_M \Omega(M, N)$ is a normalization factor, and the variance of M is

$$\sigma_M = \frac{1}{3} N \mu^2.$$

M is a sum of N independently-distributed random variables. The fact that the variance is proportional to N follows from this. The fact that it is proportional to μ^2 follows from dimensional analysis. What about the $1/3$?

Claim: $\langle m_z^2 \rangle = \frac{1}{3} \mu^2$.

$$\langle m_z^2 \rangle = \frac{1}{4\pi\mu^2} \int_{S^2} m_z^2$$

where the integration is over the two-sphere of possible values of the direction of the moment.

$$\langle m_z^2 \rangle = \frac{1}{4\pi} \int d\theta d\varphi \sin\theta (\mu^2 \cos^2\theta) = \frac{2\pi\mu^2}{4\pi} \int_{-1}^1 dx x^2 = \frac{\mu^2}{3}.$$

I defined $x \equiv \cos\theta$.

27.2 Comment on regime of validity of the answer for $p(m)$

The answer we find for the probability distribution for one moment (in part (e) of the problem) is

$$p(m) = \frac{1}{2\mu} e^{-\frac{3mM}{N\mu^2}}. \quad (27.1)$$

Notice that this is not a normalized probability distribution:

$$\int_{-\mu}^{\mu} dm p(m) = \frac{1}{2\mu} \frac{N\mu^2}{3M} \left(e^{\frac{3M}{N\mu}} - e^{-\frac{3M}{N\mu}} \right) \neq 1.$$

What went wrong?

Just like when we took the temperature too low in part d) and found $M > \mu N$, we went outside the range of validity of the given expression for $\Omega(M, N)$. This expression is a sharply-peaked gaussian in M , centered at $M = 0$ with a width of order $\sqrt{N\mu}$ – it is overwhelmingly probable that $M < \sqrt{N\mu}$ times a few. In this case, the exponent in (27.1) is small: $M \sim \sqrt{N\mu} \ll N\mu$ and we can Taylor expand and find a normalized probability distribution.

The gaussian form of the distribution for M is valid near the peak where most of the probability lies; to understand what happens in the tails, we would need more information

27.3 Comments about microcanonical assumption for continuous classical distributions

How did I pick the probability distribution for θ , the direction of the moment, above? In this case, I demanded that no direction was special and this determines the measure to be the round measure on the sphere.

More generally, the microcanonical hypothesis, *i.e.* the idea that "every accessible microstate should be equally probable" is ambiguous if our microstates are labelled by continuous variables. If we change variables in our distribution, the notion of "equally probable" actually changes. Extra information is required, such as a preferred choice of variables.

In quantum mechanics, this issue does not arise. Taking a classical limit of a quantum system gives a preferred choice of variables.

28 Heat pumps and refrigerators

Here's a strategy for thinking about problems like the last one on pset 6.

1. Draw a diagram indicating where heat can flow.
2. In each of these problems, one is asked to assume that the engine is in steady state. This means that the temperatures aren't changing, which means that the power is conserved at junctions in the diagram. It is analogous to one of the two Kirchoff's

laws, where we assume that charge does not pile up and hence the total current going in to each junction is zero.

3. In each of these problems, we are asked to assume that the engine is ideal, meaning that it is a Carnot engine, *i.e.* quasistatic and reversible. This says that the total entropy change is zero, and gives a relation between the various heats flowing and the temperatures, of the form

$$0 = \Delta S = \frac{Q_H}{T_H} - \frac{Q_C}{T_C}.$$

Note that since the engine is cyclic, ratios of powers are the same as ratios of heats:

$$\frac{T_H}{T_C} = \frac{Q_H}{Q_C} = \frac{Q_H \text{ per cycle}}{Q_C \text{ per cycle}} = \frac{P_H}{P_C}.$$

29 There is no magic about Maxwell relations

For any state variable in a system with more than one thermodynamic degree of freedom (*e.g.* temperature and volume), we can obtain a Maxwell relation by demanding that its mixed partials be equal. Any smooth function of state variables is also a state variable, so there are many such relations, not all of which are useful. Don't try to memorize them.

To try to free our understanding of Maxwell relations from the notation associated with hydrostatic systems, let's do an example of a magnet, where the thermodynamic variables are the magnetization M , temperature T , entropy S , and external magnetic field B (note that I'm calling it B instead of H to distinguish it from the enthalpy), two of which are independent.

Warmup Q: show that

$$\partial_B T|_S = -\partial_S M|_B$$

The first law applied to such a system is

$$dE = TdS + BdM.$$

Note that this equation is equivalent to the two statements

$$\partial_M E|_S = B, \quad \partial_S E|_M = T.$$

Equating the mixed partials of E gives

$$\partial_S|_M \partial_M|_S E = \boxed{\partial_S|_M B = \partial_M|_S T}.$$

This is a correct Maxwell equation, but not the one we asked about.

To change which variables appear, let's do "a Legendre transformation"; here this fancy phrase just means consider another state variable which is obtained from E by adding a product of an intensive variable and an extensive variable:

$$H \equiv E - BM$$

(it's also important that the thing we add have dimensions of energy; one way to guarantee this is if the two variables are conjugate to each other, in that one is some generalization of displacement (like V or M) and the other is the corresponding generalized force (like P or B respectively). The name H has no real significance, but I've given it this name to emphasize the similarity with the enthalpy. I chose the sign of BM to get

$$dH = TdS - MdB \quad ; \quad (29.1)$$

if we add BM , we get some more complicated expression which isn't as useful. $H = H(S, B)$ is the natural thermodynamic potential to use at fixed S, B . The equation (29.1) says that

$$\partial_S H|_B = T, \quad \partial_B H|_S = -M.$$

Equating mixed partials of H gives

$$\partial_B|_S \partial_S|_B H = \boxed{\partial_B|_S T = -\partial_S|_B M}$$

which is what we were asked to show.

There are two more nice potentials we can define for this system, which are the Helmholtz free energy $F = E - TS$ and the analog of the Gibbs free energy $G = E - TS - BM$. These give two more Maxwell relations, one of which we will need in the next problem.

30 More paramagnet

[S2003 pset 7 problem 2]

Suppose we are told that the equation of state of a paramagnet is

$$M = \frac{AB}{T - T_0}$$

for A, T_0 constants, and this equation is valid only for $T > T_0$. B here is the external magnetic field.

a) Show that the heat capacity at fixed M is independent of M at constant T , *i.e.*

$$0 = \partial_M C_M|_T.$$

By definition

$$C_M|_T \equiv \frac{dQ}{dT}|_M = \frac{TdS}{dT}|_M.$$

Since the first law for the magnet is $dE = TdS + BdM$, at fixed M all the heat goes into internal energy and this is equal to

$$C_M|_T = \frac{dE}{dT}|_M.$$

There are several ways to do this problem. Here is one which is slightly different in detail from the one in the S2003 OCW solution. The question asks about

$$\partial_M|_T C_M = \partial_M|_T \partial_T|_M E.$$

Since E is a state variable, the mixed partials have to be equal, so this is

$$\partial_M|_T C_M = \partial_T|_M \partial_M|_T E.$$

Now let's see what we know about $\partial_M|_T E$. To think about varying E at fixed T , we should rewrite the first law as follows:

$$dE = TdS + BdM = T(\partial_T|_M SdT + \partial_M|_T SdM) + BdM = T\partial_T|_M SdT + (T\partial_M|_T S + B)dM.$$

This equation says that the quantity of interest is

$$\partial_M|_T E = T\partial_M|_T S + B. \quad (30.1)$$

It also says that $\partial_T|_M E = T\partial_T|_M S$, which we knew; this is C_M .

We don't yet know $\partial_M|_T S$, though. Here we can use another Maxwell relation. Let $F = E - TS$, so $dF = -SdT + BdM$, *i.e.*

$$\partial_T F|_M = -S, \quad \partial_M F|_T = B.$$

The mixed partials of F are

$$\boxed{-\partial_M S|_T = \partial_T B|_M} \quad (30.2)$$

which is the relation we need. Then (30.1) says that

$$\partial_M|_T E = T\partial_M|_T S + B = -T\partial_T B|_M + B.$$

So far this is general. In this problem, we know $B(T, M)$ from the given equation of state

$$B(T, M) = \frac{M}{A}(T - T_0)$$

so for this system

$$\partial_M|_T E = -T\partial_T B|_M + B = -\frac{M}{A}T + \frac{M}{A}(T - T_0) = -\frac{M}{A}T_0.$$

Then

$$\partial_M|_T C_M = \partial_M|_T \partial_M|_T E = \partial_M|_T \left(-\frac{M}{A}T_0 \right) = 0.$$

b) Find $E(T, M)$ in terms of $A, T_0, C_M(T)$. Integrals of C_M and unfixed additive constants are OK.

We already figured out that

$$dE = T\partial_T|_M S dT + (T\partial_M|_T S + B) dM = C_M dT + (-T\partial_T B|_M + B) dM$$

which for this problem is

$$dE = C_M(T) dT - \frac{M}{A} T_0 dM.$$

Notice that in this differential, the coefficient of dT depends only on T and the coefficient of dM depends only on M . This makes it easy to integrate because the integral is then just the sum of a function of M and a function of T – note that this is consistent with the fact that $\partial_M|_T C_M = \partial_M|_T \partial_T|_M E = 0$. That function is

$$E(T, M) = \int^T C_M(T') dT' - \frac{M^2}{2A} T_0 + \text{const.}$$

c) Find $S(T, M)$ in terms of $A, T_0, C_M(T)$.

Similarly,

$$dS = \partial_M S|_T dM + \partial_T S|_M dT = -\partial_T|_M B dM + \frac{C_M}{T} dT = -\frac{M}{A} dM + \frac{C_M(T)}{T} dT.$$

In the second step we used again the Maxwell relation (30.2).

This is again easy to integrate:

$$S(T, M) = -\frac{M^2}{2A} + \int^T \frac{C_M(T')}{T'} dT' + \text{other const.}$$

31 Cooling a fluid: an exercise using the first law and Maxwell relations

[big Reif (*Fundamentals of Statistical and Thermal Physics*), §5.9, 5.10; Greiner, chapter 4]

A few weeks ago we talked briefly about adiabatic free expansion of a gas. Finally we can talk about another way to cool a gas. It is an opportunity to use a Maxwell relation, and a situation where a thermodynamic potential other than energy is physically relevant.

31.1 Joule-Thomson Throttling

In practice, a better way to cool fluids (which is used to liquefy gases) is the following procedure introduced by Joule and Thomson (= Lord Kelvin). This procedure also has the advantage of avoiding the intermediate unspeakable non-equilibrium situation.

The idea is to set up a steady flow of gas down an insulating pipe. In the path of the gas is a porous plug, which maintains a pressure differential. On the left, upstream side we have P_1, T_1 , on the right we have P_2 and we'd like to know T_2 .

Consider a fixed mass of gas between two points A, B along the pipe. As the chunk moves, it does work on the gas downstream of it, and it gets work done on it by the gas upstream of it. Think about the work done on this chunk of gas between the time right as it is about to enter the plug and the time right after it has all left the plug. Initially it has volume V_1 , and the gas upstream of it is doing work at pressure P_1 to decrease this volume to zero; this is an amount of work $P_1 V_1$ on the gas. The chunk itself does work against a pressure P_2 on the right to produce a final volume of V_2 , which is work $-P_2 V_2$ done ON the gas. The net work done on this chunk is

$$W = -P_2 V_2 + P_1 V_1.$$

On the other hand, no heat is absorbed because we have assumed steady state and insulating walls. So the first law says

$$0 = Q_{\text{into gas}} = \Delta U_{\text{of gas}} - W_{\text{on gas}}.$$

Note the signs (and note that Reif uses the opposite convention, using W to mean the work done *by* the gas). The change in the internal energy is

$$\Delta E = E_2 - E_1 = E(T_2, P_2) - E(T_1, P_1).$$

So

$$\Delta E = \Delta(PV)$$

or

$$E_1 + P_1V_1 = E_2 + P_2V_2 \quad \text{i.e. } H_1 = H_2$$

where $H \equiv E + PV$ is the enthalpy.

Again for an ideal gas this is boring since

$$H = E + PV = E(T) + NkT$$

depends only on T so $H(T_2) = H(T_1)$ just says the temperature is constant and nothing happens. So again we need some interactions in our gas to get an effect. The effect we want is a change in T as we change the pressure, *i.e.* a nonzero value of the “Joule-Thomson coefficient”

$$\mu \equiv \left. \frac{\partial T}{\partial P} \right|_H.$$

$\mu > 0$ ($\mu < 0$) means the gas cools (warms) as we decrease the pressure $P_2 < P_1$.

Let’s find a useful expression for μ . Recall that

$$dU = \delta Q - PdV \quad \longrightarrow \quad dH = d(E + PV) = \delta Q + VdP.$$

Thinking of $H = H(T, P)$ as function of the independent variables T, P , we have

$$dH = \left. \frac{\partial H}{\partial T} \right|_P dT + \left. \frac{\partial H}{\partial P} \right|_T dP.$$

At constant enthalpy, we set this to zero and learn

$$\mu = \left. \frac{\partial T}{\partial P} \right|_H = - \frac{\left. \frac{\partial H}{\partial P} \right|_T}{\left. \frac{\partial H}{\partial T} \right|_P}.$$

We’ve seen in lecture that the thing in the denominator is

$$C_P = \left. \frac{\partial H}{\partial T} \right|_P$$

the specific heat at constant pressure.

Now let’s work on this some more to relate it to familiar quantities. The first law $dE = TdS - PdV$ and the definition of enthalpy $H = E + PV$ imply that

$$dH = TdS + VdP = T(\partial_T S|_P dT + \partial_P S|_T dP) + VdP = C_P dT + (T\partial_P S|_T + V) dP.$$

Setting $dH = 0$, the Joule-Thomson coefficient is then

$$\mu = \left. \frac{\partial T}{\partial P} \right|_H = - \frac{T\partial_P S|_T + V}{C_P}.$$

Now we can use a Maxwell relation.

$$G = E - TS + PV = H - TS, \quad dG = -SdT + VdP = \partial_T G|_V dT + \partial_P G|_T dP$$

so $S = -\partial_T G|_V$, $V = \partial_P G|_T$. So equality of mixed partials, *i.e.* the fact that G is a state function, implies

$$\boxed{\partial_P S|_T = \partial_T V|_P} = V\alpha$$

where α is the coefficient of expansion (the second equality here is actually its definition).

So

$$\mu = -\frac{T\partial_P S|_T + V}{C_P} = \frac{V}{C_P}(T\alpha - 1).$$

So Joule-Thomson throttling will cool the gas when $\mu > 0$ which happens when $\alpha > 1/T$, *i.e.* $\{\alpha = 1/T\}$ gives the inversion curve.

32 Warmup problem: defects in a solid

[based on S2003 pset 7 number 5]

Consider N independent but distinguishable several-state systems. Let each system have $l_0 + l_1$ states such that $\epsilon_1 = \epsilon_2 = \dots = \epsilon_{l_0} = 0$ and $\epsilon_{l_0+1} = \epsilon_{l_0+2} = \dots = \epsilon_{l_0+l_1} = \Delta$ for some parameter Δ .

(a) Find the partition function for one such system in equilibrium at temperature T .

$$Z_1 = \sum_{\alpha=1}^{l_0+l_1} e^{-\beta\epsilon_\alpha} = l_0 + l_1 e^{-\beta\Delta}$$

where $\beta \equiv \frac{1}{k_B T}$.

(b) Find the partition function $Z(T, N)$ of N of them.

$$Z = Z_1^N.$$

The fact that they are distinguishable means no $N!$.

(c) Find the entropy.

$$F = -k_B T \ln Z = -k_B T N \ln Z_1$$

$$S = -\partial_T F = k_B N \ln Z_1 + \frac{k_B T N}{Z_1} l_1 \Delta (-\partial_T \beta) = k_B N \ln Z_1 + \frac{k_B T N}{Z_1} l_1 \Delta \frac{1}{k_B T^2}.$$

(d) What is the contribution of these things to the entropy at high temperature, $k_B T \gg \Delta$?

At high temperature, the Boltzmann suppression factor which prevents very high-energy stuff from happening goes to unity:

$$e^{-\beta \Delta} \xrightarrow{k_B T \gg \Delta} 1$$

This says that all states are equally likely, independent of their energy, and the partition function just counts the number of states.

$$Z_1 \xrightarrow{k_B T \gg \Delta} l_0 + l_1$$

The average energy of one subsystem is

$$\langle \epsilon \rangle = \sum_{\alpha} \epsilon_{\alpha} P_{\alpha} \xrightarrow{k_B T \gg \Delta} \frac{l_1 \Delta}{l_0 + l_1}.$$

The average energy of N independent subsystems is N times as big,

$$\langle E \rangle = \frac{N l_1 \Delta}{l_0 + l_1}.$$

We can reproduce this answer by finding

$$E = F + TS = \frac{k_B T^2 N}{Z_1} l_1 \Delta \xrightarrow{k_B T \gg \Delta} \frac{N l_1 \Delta}{l_0 + l_1}.$$

33 Information and entropy

[from Blundell]

Consider the following three statements:

1. Isaac Newton's birthday occurs on a particular day of the year.
2. Isaac Newton's birthday occurs in the first half of the year.
3. Isaac Newton's birthday occurs on the 25th of a month.

We can agree that they are ordered from least informative to most informative. A way to quantify the information content of a statement is to notice that a statement is less informative if the probability that it is true without any prior information is larger:

$$P_1 = 1, \quad P_2 = \frac{1}{2}, \quad P_3 = \frac{12}{365}.$$

To figure out how to make the relationship between this prior probability and information content more precise let's consider how to combine independent statements. Statements 2 and 3 can be independently true or false; we would like to say that the information content for knowing *both* is the sum of the individual information contents. The probabilities on the other hand multiply:

$$P_{2 \text{ and } 3} = P_2 \cdot P_3 = \frac{6}{365}.$$

A definition of the information content which satisfies this rule of combination was introduced by Shannon:

$$Q \equiv -k \log P.$$

k is a positive constant; if $k = 1$ and $\log = \log_2$, information is measured in bits, while if $k = k_B$ and $\log = \ln$, information has the same units as thermodynamic entropy.

Given a set of statements i to which we can assign probabilities P_i , we have $Q_i = -k \log P_i$. Define the *Shannon entropy* to be the average information content over the distribution,

$$S \equiv \langle Q \rangle = \sum_i P_i Q_i = -k \sum_i P_i \log P_i.$$

This quantifies how much information we would gain upon measuring the observable i , or alternatively how much uncertainty we have before we measure it.

We can associate such an entropy with any probability distribution. For example, suppose there are two outcomes $i = 0, 1$ and $P_0 \equiv p, P_1 = 1 - p, p \in [0, 1]$, like one step of a 1d random walk (such a thing is called a Bernoulli process). Then

$$S = -k(p \log p + (1 - p) \log(1 - p)).$$

This function vanishes at the two extremes $p = 0, 1$ where we are sure about the outcome of measuring i . It has a maximum when $p = \frac{1}{2}$ where we are maximally uncertain about the outcome.

33.1 Entropy and probability

Now we will try to connect to the derivation of the canonical ensemble in problem 4 of pset 7.

Suppose we have a system which can be in one of several macrostates labelled by a variable $i = 1, \dots$, which is easy to measure. Let n_i be the number of microstates associated with the macrostate i ; we should imagine that these microstates are impossible for us to distinguish. An example to think about is rolling a die where the number

that ends up on top is $i = 1 \dots 6$; n_i the number of ways the die could have tumbled and ended up with i on top. Let $N \equiv \sum_i n_i$ be the total number of microstates. Then we can assign probabilities by counting

$$P_i = \frac{n_i}{N}$$

and $\sum_i P_i = 1$ is automatic.

The total microcanonical entropy if we know nothing about the state of this system is

$$S_{\text{tot}} = k_B \ln N.$$

Because of our ability to distinguish the macrostates i , we can decompose this into two parts:

$$S_{\text{tot}} = S + S_{\text{micro}}$$

where S is the thermodynamic entropy associated with not knowing which macrostate i the system is in, while

$$S_{\text{micro}} \equiv \sum_i P_i S_i = \sum_i P_i k_B \log n_i$$

is the entropy associated with not knowing the microstate – $S_i \equiv k_B \log n_i$ is just the usual microcanonical entropy of the macrostate i . Rearranging this gives an expression for the physical entropy

$$S = S_{\text{tot}} - S_{\text{micro}} = k_B \left(\log N - \sum_i P_i \log n_i \right) = k_B \sum_i P_i (\log N - \log n_i) = -k_B \sum_i P_i \log P_i.$$

In the third step we inserted $1 = \sum_i P_i$ in the first term. This is just the Shannon entropy of the probability distribution P_i . This expression for the entropy is due to Gibbs.

Now suppose that we knew nothing at all about physics but were given the following assignment: assign probabilities to some set of outcomes labelled by a variable i , consistent with the constraints

$$(1) \quad \sum_i P_i = 1$$

$$(2) \quad \sum_i P_i E_i = E.$$

The second constraint says that we know the average energy. There are in general many ways to assign probabilities satisfying these constraints; our assignment is to do this in the way that *adds the fewest unwarranted assumptions* about the system, *i.e.*

that adds the least amount of information that we don't actually have, *i.e.* that is least biased.

From the discussion of the previous subsection, we see that the answer to this problem is to choose the set of P_i which maximizes the Shannon entropy, subject to the constraints (1) and (2). This will be the probability distribution with the least information (by Shannon's definition) consistent with the available data.

To implement this, we can use Lagrange multipliers α, β and extremize

$$I(P_i, \alpha, \beta) \equiv S(P) - \alpha \left(\sum_i P_i - 1 \right) - \beta \left(\sum_i P_i E_i - E \right).$$

Varying with respect to α imposes probability normalization, varying with respect to β inputs the average energy. Varying with respect to P_i gives

$$0 = -\ln P_i + 1 - \alpha - \beta E_i$$

which implies

$$P_i = e^{-1-\alpha} e^{-\beta E_i}.$$

Demanding that this be normalized gives

$$P_i = \frac{e^{-\beta E_i}}{Z}.$$

So we see that in fact the Boltzmann distribution is much more inevitable than even we would think from pset 7 problem 4. It does not rely on any metaphysics about equal probabilities for microstates. It is merely the least-biased distribution which we can assign. This point of view was forcefully advocated by E.T. Jaynes. You may enjoy his original paper on the subject, [here](#).

34 Warmup problem: Schottky anomaly

Consider a two-state system. The two states are called $+$ and $-$ and have energies $\epsilon_{\pm} = \pm\Delta/2$. Compute and sketch the specific heat as a function of temperature. Compute and sketch the specific heat as a function of temperature for a system comprised of N distinguishable copies of such a system.

$$Z_1 = \sum_r e^{-\beta\epsilon_r} = e^{+\beta\Delta/2} + e^{-\beta\Delta/2} = 2 \cosh\left(\frac{\beta\Delta}{2}\right).$$

$$\langle\epsilon\rangle = -\partial_{\beta} \ln Z = -\frac{\Delta}{2} \tanh\left(\frac{\beta\Delta}{2}\right).$$

$$C_V = \partial_T \langle \epsilon \rangle = k_B \left(\frac{\beta \Delta}{2} \right)^2 \operatorname{sech} \left(\frac{\beta \Delta}{2} \right).$$

The specific heat goes to zero at large T and at small T . At large T this is because both states become equally probable; varying a big temperature doesn't change anything. At small T this is because the system is frozen into its groundstate, which is the state called $-$. The big bump at $k_B T \sim \Delta$ is called the ‘‘Schottky anomaly’’ and is where all the action is. When the thermal energy $k_B T$ is of order the level spacing Δ then thermal excitations have an order-one probability of exciting transitions; changing the temperature makes a big difference for the average energy.

For N distinguishable copies, $Z_N = Z_1^N$, and $F_N = -k_B T \ln Z_N = -k_B N T \ln Z_1$. The specific heat and the energy are just $E_N = N \langle \epsilon \rangle$, N times as big.

35 Comment on derivation of canonical ensemble from microcanonical ensemble

In lecture Krishna derived the canonical ensemble by considering a subsystem (1) of a large, but isolated system (1+2). The probability that subsystem (1) is in some microstate A is the fraction of microstates of the whole system where this is the case:

$$p((1) \text{ is in a state } A \text{ with energy } E_1) = \frac{\Omega'}{\Omega_{(1+2)}(E_{\text{TOT}})} = \frac{1 \times \Omega_{(2)}(E_{\text{TOT}} - E_1)}{\Omega_{(1+2)}(E_{\text{TOT}})}.$$

Then we Taylor expanded the log of p (the log of any function is more smoothly varying than the function itself and so its Taylor expansion is a better approximation):

$$\Omega_2(E_{\text{TOT}} - E_1) \approx e^{-\beta E_1} e^{-S_2(E_{\text{TOT}})/k_B}$$

where

$$\beta = \frac{\partial S_2(E_{\text{TOT}})}{\partial E_{\text{TOT}}} = \frac{1}{k_B T}$$

makes its first appearance here. This gives us an expression for the probability proportional to the Boltzmann factor

$$p((1) \text{ is in state } A) = \frac{e^{-\beta E_1}}{Z}$$

with

$$\frac{1}{Z} = e^{-(S_2(E_{\text{TOT}}) + S_{1+2}(E_{\text{TOT}}))/k_B}.$$

So far, this is just a review what happened in lecture. Next we made the assumption that (1) is *also* macroscopic. It is the role of this assumption that I want to clarify. This led us to write

$$S_{\text{TOT}}(E_{\text{TOT}}) = S_1(\langle E_1 \rangle) + S_2(\langle E_2 \rangle);$$

I want to explain in more detail how this follows. It is an opportunity to explain a useful technique for approximating a certain class of integrals.

First, we have

$$e^{S_{\text{TOT}}(E_{\text{TOT}})} = \Omega_{1+2}(E_{\text{TOT}}) = \int_0^{E_{\text{TOT}}} dE_1 \Omega_1(E_1) \Omega_2(E_{\text{TOT}} - E_1).$$

The first step here is the microcanonical definition of entropy. The second step follows because we assume that systems (1) and (2) are independent but able to exchange energy. This means that the energy of system (1) is not fixed, and we must sum over its values (if we were treating the system quantum mechanically, it would really be a sum not an integral; let's leave this refinement for another time). This is

$$\int_0^{E_{\text{TOT}}} dE_1 \Omega_1(E_1) \Omega_2(E_{\text{TOT}} - E_1) = \int_0^{E_{\text{TOT}}} dE_1 e^{(S_1(E_1) + S_2(E_{\text{TOT}} - E_1))/k_B}.$$

Now, if (1) and (2) are *both* macroscopic, then each of these entropies in the exponent of the integrand are of the form $S(E) = Ns(E)$ where N is a big number related to the number of degrees of freedom, and s is some intensive function; this follows because the entropy is extensive. So our integral is something like

$$I = \int_0^{E_{\text{TOT}}} dE_1 e^{Ns(E_1)}. \tag{35.1}$$

Aside: The Saddle Point Method

Consider an integral of the form above

$$I = \int dx e^{-Nf(x)}$$

where $N \gg 1$ is a big number and $f(x)$ is a smooth function. As you can see from the example in the figure (where N is only 10), $e^{-Nf(x)}$ is hugely peaked around the absolute minimum of $f(x)$, which I'll call x_0 . We can get a good approximation to the integral by just considering a region near $x = x_0$, and Taylor expanding $f(x)$ about this point:

$$f(x) = f(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) + \dots$$

where there's no linear term since x_0 is a critical point, and we assume a minimum $f''(x_0) > 0$. It is also important that $x = x_0$ is in the range of integration. Then

$$\begin{aligned} I &= \int dx e^{-Nf(x)} \approx \int dx e^{-N(f(x_0) + \frac{1}{2}(x-x_0)^2 f''(x_0) + \dots)} \\ &= e^{-Nf(x_0)} \int dy e^{-\frac{N}{2} f''(x_0) y^2 + \dots} \approx e^{-Nf(x_0)} \sqrt{\frac{2\pi}{N f''(x_0)}}. \end{aligned}$$

The important bit is that the integral is well-approximated by $e^{-Nf(x_0)}$, *i.e.* just plugging in the value at the critical point. (Actually, for values of N as small as I've chosen in the example, the bit with the f'' is important for numerical accuracy; for the example in the figure, including this factor, the saddle point method gives $\int_{-2}^2 dx e^{-Nf(x)} = 206.7$, while numerical integration gives 209.3. Without the gaussian integral correction, saddle point gives 816.6. For values of $N \sim 10^{24}$ we can just keep the leading term.)

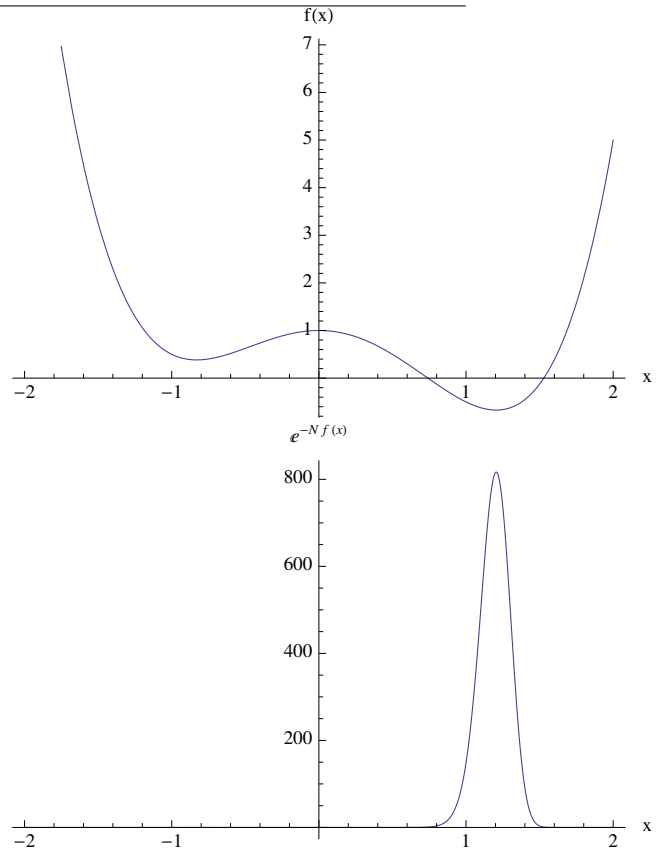


Figure 18: top: $f(x) = (x^2 - 1) - \frac{1}{2}x^3$
bottom: $e^{-Nf(x)}$ with $N = 10$.

Applying this technique to the integral above (35.1) the condition for the saddle point is

$$0 = \frac{\partial s(E_1)}{\partial E_1} = \frac{\partial}{\partial E_1} (S_1(E_1) + S_2(E_{\text{TOT}} - E_1)) = \frac{1}{T_1} - \frac{1}{T_2}.$$

The saddle point condition says that the temperatures of the two systems should be equal. This determines the value of E_1 to be the thermal equilibrium value at the temperature of the bath $\frac{1}{T} = \partial_{E_2} S_2$. Plugging in the saddle point approximation to the E_1 integral then, we have

$$\Omega_{1+2}(E_{\text{TOT}}) = e^{(S_1(\langle E_1 \rangle) + S_2(\langle E_2 \rangle))/k_B}$$

as desired.

36 Quantum ideal gas of distinguishable particles

Consider a system where the states are labelled by three positive integers,

$$n_x, n_y, n_z = 1, 2, 3, 4, \dots, \infty$$

and the energies of the states are

$$\epsilon_{n_x, n_y, n_z} = \epsilon_0 (n_x^2 + n_y^2 + n_z^2).$$

This is realized by a particle of mass m in a cubical box of side length L ; in that case $\epsilon_0 = \frac{h^2}{8mL^2}$. For our purposes, this merely a cultural remark.

The partition function for such a system in thermal equilibrium at temperature T is

$$Z_1 = \sum_{n_x=1}^{\infty} \sum_{n_y=1}^{\infty} \sum_{n_z=1}^{\infty} e^{-\beta\epsilon_0(n_x^2+n_y^2+n_z^2)} = \left(\sum_{n=1}^{\infty} e^{-\beta\epsilon_0 n^2} \right)^3.$$

The sum here is not so simple; in fact it is a function called a Jacobi theta function. We will make progress by considering the high-temperature and low-temperature limits.

36.1 High temperature

The only energy scale in the system is ϵ_0 , so by high temperature I mean $k_B T \gg \epsilon_0$. In this limit, the sum over ns is dominated by large n : the particles are in highly-excited states. For this part of the sum, the values of successive terms in the sum are not very different, and we can make a good approximation

$$\sum_{n=1}^{\infty} e^{-\beta\epsilon_0 n^2} \approx \int_0^{\infty} dn e^{-\beta\epsilon_0 n^2} = \frac{1}{2} \int_{-\infty}^{\infty} dn e^{-\beta\epsilon_0 n^2} = \frac{1}{2} \sqrt{\frac{\pi}{\beta\epsilon_0}}.$$

So we have

$$Z_1 \stackrel{\beta\epsilon_0 \ll 1}{\approx} \frac{V}{h^3} \left(\frac{2\pi m}{\beta} \right)^{3/2}$$

which is in fact the classical ideal gas partition function,

$$Z_1 (\beta\epsilon_0 \ll 1) \approx \frac{V}{h^3} \left(\frac{2\pi m}{\beta} \right)^{3/2} = \int \frac{d^3p d^3x}{h^3} e^{-\beta p^2/2m}.$$

36.2 Aside about variance of energy

We know that $\langle \epsilon \rangle = -\partial_\beta \ln Z$. I claim that the quantity $(-\partial_\beta)^2 \ln Z$ is also interesting for several reasons.

On the one hand, it is proportional to the specific heat:

$$(-\partial_\beta)^2 \ln Z = -\partial_\beta \langle \epsilon \rangle = -k_B T^2 C_V.$$

On the other hand,

$$(-\partial_\beta)^2 \ln Z = -\partial_\beta \left(\frac{\sum_r \epsilon_r e^{-\beta \epsilon_r}}{Z} \right) = \frac{\sum_r \epsilon_r^2 e^{-\beta \epsilon_r}}{Z} - \left(\frac{\sum_r \epsilon_r e^{-\beta \epsilon_r}}{Z} \right)^2 = \langle \epsilon^2 \rangle - \langle \epsilon \rangle^2 = \Delta E^2$$

it is the variance of the energy.

Suppose we have a system which is N distinguishable copies of a smaller system (*e.g.* N distinguishable non-interacting quantum particles). Then

$$Z = Z_1^N, \quad \ln Z = N \ln Z_1.$$

Using our expressions above and differentiating $\ln Z$,

$$\langle E \rangle = N \langle \epsilon \rangle, \quad \Delta E^2 = N \Delta \epsilon^2.$$

So the relative fluctuations of the energy of the big system are

$$\frac{\langle E \rangle}{\Delta E} = \frac{N \langle \epsilon \rangle}{\sqrt{N \Delta \epsilon^2}} \propto \frac{1}{\sqrt{N}} \xrightarrow{N \rightarrow \infty} 0.$$

This is what makes it OK to just plug in the equilibrium value of the energy and not worry about statistical fluctuations for macroscopic systems⁶ in thermal equilibrium.

⁶at least ones composed of non-interacting subsystems

36.3 Low temperature

When the thermal energy $k_B T$ is of order the level spacing ϵ_0 , we can no longer approximate the sum as an integral, because successive terms will be quite different. Let's think about extreme low temperatures $\beta\epsilon_0 \rightarrow \infty$.

What is the average energy of the quantum particle in a box when the temperature is zero? At $T = 0$ the system is in its groundstate. In this particular case, this is $n_x = n_y = n_z = 1$, so $\epsilon_{gs} = 3\epsilon_0$. Therefore

$$\langle \epsilon \rangle (T = 0) = 3\epsilon_0.$$

Let's figure out the leading correction to this in $\frac{\epsilon_0}{k_B T}$. For very small T , the partition function is

$$Z_1 = e^{-\beta 3\epsilon_0} + 3e^{-\beta 6\epsilon_0} + \dots$$

where the first term is the groundstate contribution, the second term is from the first excited state, and higher excited states give even more exponentially-suppressed contributions. The first excited state is when two of the n s are 1 and one is 2; the 3 in the second term is from the three ways to do this; the energy of the first excited state is

$$\epsilon_{\text{1st excited state}} = \epsilon_0 (1 + 1 + 2^2) = 6\epsilon_0.$$

This expression for the partition function is simply a Taylor expansion in the small quantity $e^{-\beta\epsilon_0}$. The average energy is then

$$\begin{aligned} \langle \epsilon \rangle (k_B T \ll \epsilon_0) &\approx 3\epsilon_0 P(\text{ground state}) + 6\epsilon_0 P(\text{1st excited state}) + \dots \\ &= 3\epsilon_0 \frac{e^{-\beta 3\epsilon_0}}{e^{-\beta 3\epsilon_0} + \text{small}} + 6\epsilon_0 \frac{3e^{-\beta 6\epsilon_0}}{e^{-\beta 3\epsilon_0} + \text{small}} + \text{small} \\ \langle \epsilon \rangle (k_B T \ll \epsilon_0) &\approx 3\epsilon_0 + 18\epsilon_0 e^{-\beta 3\epsilon_0} + \text{small}. \end{aligned}$$

The role of the fancy theta function is to interpolate between this answer at low temperatures and the ideal gas $\frac{3}{2}k_B T$ behavior at high temperatures.

37 Entropy of blackbody radiation and a puzzle

[based on Baierlein section 6.4]

Here is an apparent paradox. 3.5 billion years ago the only life on earth was a bunch of bacteria. Surely there are more microstates of the atoms making up our

planet where there are only bacteria than where there are humans and civilization and physicists. So it seems the entropy of our planet has decreased,

$$\Delta S_{\text{earth}} = S_{\text{earth}}(\text{now}) - S_{\text{earth}}(\text{bacteria times}) < 0.$$

Does this violate the second law of thermodynamics? No, not obviously, because the Earth is not an isolated system.

In particular, energy is being pumped in via light from the sun at temperature T_S , and pumped out via thermal radiation of the earth itself, which we can very roughly regard for this purpose as being in thermal equilibrium at temperature T_E . Demanding that the situation is static tells us that the energy per time going in had better equal the energy per time going out,

$$0 = aT_S^4 - bT_E^4$$

and this relates the temperature of the sun and of the earth via some geometric factors (which I've called a, b here) as on pset 9.

What is the entropy budget?

37.1 Entropy of blackbody radiation

The entropy of a volume V of radiation at temperature T is

$$S_{\text{rad}}(T, V).$$

We figured this out in lecture using the microcanonical ensemble, but here is a way to remember the answer without calculation.

The entropy divided by k_B is dimensionless; it's the log of a number. And it depends only on T, V and a bunch of constants of nature, specifically \hbar, c, k_B . We can choose to work in units where these constants are equal to one. This means that they can't be combined to make dimensionless combinations on their own. So we have to find a dimensionless combination x of T and V using these constants, and then the entropy can be any function of x . I claim that a good choice of x is

$$x = k_B T \frac{V^{1/3}}{\hbar c}.$$

The first factor $k_B T$ is an energy. Evidence that

$$\left[\frac{\hbar c}{V^{1/3}} \right] = \text{energy}$$

is that $\frac{\hbar}{v^{1/3}}$ has units of momentum, and momentum times velocity is an energy (recall $E = pc$ for photons). So the entropy only depends on V and T in the combination VT^3 :

$$S_{rad}(T, V) = S(VT^3).$$

But we know that entropy is extensive, and the only extensive thing in sight is the volume, so

$$S \propto V$$

which implies

$$S = \text{cst}VT^3$$

for some constant cst.

This tells us the shape of adiabats for thermal radiation just using dimensional analysis and extensivity of the entropy.

37.2 The history of the earth is consistent with thermodynamics

The entropy change of the earth from the thermal photons arriving and leaving is then of the form

$$aT_S^3 - bT_E^3.$$

This is a negative number: the entropy of a chunk of thermal radiation with energy E at temperature T_S is *smaller* than a chunk with the same energy but a lower temperature T_E . A way to understand this is that the entropy of such a chunk of radiation is proportional to the *number of photons*. I'll justify this claim further in the next recitation. Basically, the microstates of the radiation are specified by what the photons are doing; a larger entropy means we have to specify the data for more photons.

This explains the inequality above: Higher-temperature radiation has a higher peak frequency; since the energy of a photon is proportional to its frequency, each of the photons having more energy (with fixed total energy) means fewer photons.

38 “Quietly emitting thermal radiation” ?

I'd like to pose a puzzle raised by an offhand phrase used by Krishna in lecture last week. He described a hot object as sitting somewhere “quietly emitting thermal radiation”. This raises the question: if the body is immersed in air, does it not also excite thermal sound radiation? Can you hear a hot object?

Here's a wrong calculation that suggests that a hot object should create a deafening roar of thermal acoustic radiation. Two differences between light and sound (in a fluid like air) are:

- (1) light has two (transverse) polarizations, while sound in air has only one (longitudinal) mode per wavevector, as I'll describe further below.
- (2) Light moves at the speed of light $c = 3 \cdot 10^8 m/s$ while sound in air moves at the speed of sound in air, $c_s \sim 3 \cdot 10^2 m/s$. (There is a third most important difference that I'll elide for now and let you think about it.) Recall that the energy flux in EM radiation from a body of temperature T is

$$\frac{P_{\text{light}}}{A} = \frac{4\sigma_{SB}}{c} T^4 = \frac{\pi^2 k_B^4}{15c^3 \hbar^3} T^4$$

Making the replacements above, the energy flux in sound radiation from a body of temperature T might be

$$\frac{P_{\text{sound}}}{A} \stackrel{?}{=} \frac{1}{2} \frac{\pi^2 k_B^4}{15c_s^3 \hbar^3} T^4 = \frac{1}{2} \left(\frac{c}{c_s} \right)^3 \frac{P_{\text{light}}}{A} \sim 10^{18} \frac{P_{\text{light}}}{A}.$$

This is wrong. Hint: what would be the peak wavelength of the acoustic radiation?

38.1 Hydrodynamic description

As an example let's talk more about how to describe fluctuations of the density of a fluid like air. The field variable will be

$$\rho(t, \vec{r}) \equiv \text{the density of air} - \frac{N}{V};$$

where I've subtracted off the average density, $\rho_{\text{average}} = N/V$. **If** the density doesn't vary too rapidly in time and space, the energy of any configuration is

$$H[\rho] = \int_{\text{space}} d^3r \kappa \left(\dot{\rho}^2 + c_s^2 \left(\vec{\nabla} \rho \right)^2 \right) + \dots$$

Here κ is a constant which gets the dimensions right; c_s will be the speed of sound. The ... represent terms that we will neglect; this includes terms more nonlinear in ρ , like ρ^3 which we can neglect if the deviations from the mean density are small. It also includes terms involving more derivatives of ρ which we neglect because of our assumption that variations are slow.

Since the energy density here is translation invariant, it is useful to do Fourier decomposition. This means we expand our function $\rho(\vec{r})$ in a basis of Fourier modes:

$$\rho(\vec{r}, t) = \sum_{\vec{k}} a_{\vec{k}}(t) e^{i\vec{k} \cdot \vec{r}}. \quad (38.1)$$

Here a is for ‘amplitude’. The density of air is a real variable; the RHS here is not manifestly so. If we find a complex solution of the equations of motion, its real part will be a solution too since the equation will be linear (this will become even more evident below). Plugging this into the energy expression gives:

$$\begin{aligned} H[\rho = \sum a_k e^{ikr}] &= \int_{\text{space}} d^3r \kappa \sum_{k_1, k_2} \left(\dot{a}_{k_1} \dot{a}_{k_2} + c_s^2 (i\vec{k}_1 \cdot i\vec{k}_2) a_{k_1} a_{k_2} \right) e^{i(\vec{k}_1 + \vec{k}_2) \cdot \vec{r}} \\ &= \tilde{\kappa} \sum_k \left(\dot{a}_k^2 + c_s^2 \vec{k}^2 a_k^2 \right) \end{aligned}$$

In the last step we used

$$\int dx e^{ikx} = 2\pi\delta(k)$$

three times.

This is sum of harmonic oscillator hamiltonians, one for each value of k , whose oscillation frequency is

$$\omega_{\vec{k}} = c_s |\vec{k}|.$$

38.2 Density of states

Above I’ve written \sum_k without being very specific about what the sum was over. In detail, this depends on choosing boundary conditions. My job here is to convince you that most interesting questions about thermodynamics do not depend on this choice.

The simplest thing, which we’ve done for the EM field in lecture, is to imagine that the box has walls and the field has to vanish at the walls, say

$$\rho(x = 0, y, z, t) = 0 \text{ and } \rho(x = L, y, z, t) = 0 \text{ for all } z, y, t.$$

For the moment let’s just completely ignore y, z . In this case we are forced to choose the phases in our Fourier decomposition so that each mode vanishes at $x = 0$ (note that this is true mode by mode because the condition holds for all time and the modes have different frequencies):

$$\rho = \sum_k a_k \sin k_x x.$$

Note that only $k_x > 0$ labels different states now because $\sin(-k_x x) = -\sin k_x x$ is just the same mode with a different value of a_k . Then the condition at $x = L$ constrains the wavenumbers:

$$0 = \sin k_x L \leftrightarrow k_x L = n\pi, \quad n = 1, 2, 3, \dots$$

The same thing happens for y, z :

$$\rho = \sum_{\vec{k}} a_{\vec{k}} \sin k_x x \sin k_y y \sin k_z z,$$

with

$$k_{x,y,z} = \frac{n_{x,y,z}\pi}{L}, \quad n_{x,y,z} = 1, 2, 3, \dots$$

If we are allowed to squint at the space of wavevectors, we can count them by doing integrals. If we ask questions where only a few wavenumbers contribute, this analysis will not be sufficient. This will happen for example if the temperature is too small, where the meaning too small we can figure out by dimensional analysis: we mean compared to an energy scale set by the length of the box. As is the discussion of entropy of blackbody radiation to get an energy using a length and fundamental constants, we take $\frac{\hbar c}{L}$. So ‘low temperature’ here means

$$k_B T < \frac{\hbar c}{L}.$$

If the temperature is above this value, the coarse-graining procedure here is fine, and the results are independent of boundary conditions, as we’ll see in a moment.

So

$$\begin{aligned} D(k)dk &\equiv \text{the number of modes whose wavenumber has a magnitude in the range } [k, k + dk] \\ &= \frac{\text{volume of one octant of a shell of radius } k \text{ and thickness } dk}{\text{volume of } k\text{-space occupied by each mode}} \\ &= \frac{\frac{1}{8}4\pi k^2 dk}{(\pi/L)^3}. \end{aligned}$$

Note that so far what we’ve done to determine $D(k)$ depends not at all on how the energy modes depends on k (except to figure out the range of temperatures where we can squint at the wavenumber space). The relation between frequency and wavenumber is called the ‘dispersion relation’ of the modes.

38.3 Enter the dispersion relation

The density of states in energy is defined by

$$D(\omega)d\omega = D(k)dk.$$

Multiplying by the window of energy it's exactly the same quantity we defined above, but the size of the energy window given a fixed wavenumber window is determined by the dispersion relation. So:

$$D(\omega) = D(k) \frac{1}{\left| \frac{d\omega}{dk} \right|}.$$

Here $D(k)$ we computed above (*e.g.* for longitudinal modes in 3d), and the derivative in the denominator is easy to find given $\omega(k)$.

38.4 Periodic boundary conditions

Suppose instead of hard walls of the box we just make space periodic in each direction with period L . This is a box with no walls (!). What I mean is demand that

$$\rho(x + L, y, z, t) = \rho(x, y, z, t) \text{ for all } x, y, z, t.$$

and similarly for shifting y, z . Each direction of space is a circle (the three directions together comprise a three-torus, denoted T^3); going too far in the $+x$ direction you come back to where you started from the other side. This is a very convenient theoretical device but a bit hard to implement in the lab.

This condition imposes the following relation on our Fourier decomposition (38.1):

$$e^{ik_x(x+L)} = e^{ik_x x}$$

i.e.

$$e^{ik_x L} = 1$$

i.e.

$$k_x L = 2\pi n_x, \quad n_x \in \mathbb{Z}.$$

Note two things about this:

- (1) the spacing between allowed wavenumbers is twice as big as before.
- (2) negative wavenumbers are no longer redundant; if you like we are now allowed to have both sines and cosines in our series, so there are twice as many modes for each fixed $|k_x|$.

This happens in each dimension. The volume in k -space occupied by each mode is 2^3 times as big, but the total volume of the set of modes in the shell of radius k and thickness dk is 2^3 times as big because we keep the whole shell not just the first octant. The density of states is therefore the same as it was with Dirichlet boundary conditions at the walls.

In retrospect at least it's obvious that this had to be so for large wavenumbers: imagine you are a small creature (of size $\ll L$) far from the walls of the box. Surely

you don't care about what the boundary conditions are. The density of modes of the field that you see should be independent of the choice of boundary conditions.

39 Warmup problem: number of photons

Show that the number of photons in a region of empty space of volume V , held at temperature T is

$$\mathcal{N}(V, T) \equiv \sum_{\text{modes}, \vec{k}} \langle n_{\vec{k}} \rangle = \mathcal{I} \frac{1}{\pi^2} V \left(\frac{k_B T}{\hbar c} \right)^3$$

where

$$\mathcal{I} \equiv \int_0^\infty \frac{x^2 dx}{e^x - 1} = 2.404.$$

Note that we are thinking of the average harmonic-oscillator occupation number $\langle n_{\vec{k}} \rangle$ as the number of photons with momentum \vec{k} .

$$\begin{aligned} \mathcal{N}(V, T) &= \int_0^\infty d\omega D(\omega) \langle n_\omega \rangle \\ \langle n_\omega \rangle &= \frac{1}{\hbar\omega} \left\langle \epsilon_\omega - \frac{1}{2} \hbar\omega \right\rangle = \sum_{n=0}^\infty p_n n \\ &= \sum_n \frac{n e^{-\beta\hbar\omega(n+\frac{1}{2})}}{Z} = \partial_a \left(\frac{1}{1 - e^{-a}} \right) \Big|_{a=\beta\hbar\omega} e^{-\beta\hbar\omega/2} \left(\frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} \right)^{-1} = \frac{1}{e^{\beta\hbar\omega} - 1}. \end{aligned}$$

$$\mathcal{N}(V, T) = \int_0^\infty d\omega D(\omega) \frac{1}{e^{\beta\hbar\omega} - 1} = \int_0^\infty 2 \left(\frac{1}{2} \frac{L^3 \omega^2}{\pi^2 c^3} \right) \frac{1}{e^{\beta\hbar\omega} - 1} = \frac{V}{\pi^2} \left(\frac{k_B T}{\hbar c} \right)^3 \mathcal{I} = \text{stuff} V T^3,$$

where the stuff is numbers and fundamental constants.

Recall that the entropy is of the same form. In fact:

$$S = k_B 3.602 \mathcal{N}.$$

39.1 A better proof relating entropy to occupation numbers

For a single harmonic oscillator (think of it as a particular mode of a (bosonic) field with frequency ω), show that:

$$S = k_B ((\langle n \rangle + 1) \ln (\langle n \rangle + 1) - \langle n \rangle \ln \langle n \rangle)$$

with

$$\langle n \rangle = \frac{1}{e^{\beta\hbar\omega} - 1}$$

is the occupation number expectation.

$$F_{SHO} = -kT \ln Z = -kT \ln \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}} = \frac{1}{2}\hbar\omega - kT \ln \frac{1}{1 - e^{-\beta\hbar\omega}}.$$

The first term is independent of T so won't matter when we differentiate with respect to T to get

$$\frac{S}{k_B} = -\frac{1}{k_B} \partial_T F|_V = \ln \frac{1}{1 - e^{-\beta\hbar\omega}} - \beta \partial_\beta \ln \frac{1}{1 - e^{-\beta\hbar\omega}} = \ln \frac{1}{1 - e^{-\beta\hbar\omega}} - (-1)^3 \frac{\beta\hbar\omega e^{-\beta\hbar\omega}}{1 - e^{-\beta\hbar\omega}}.$$

Now note that $\langle n \rangle$ has some weird but elementary properties:

$$\langle n \rangle + 1 = \langle n + 1 \rangle = \frac{1}{e^{-\beta\hbar\omega}}.$$

$$\frac{\langle n \rangle + 1}{\langle n \rangle} = e^{\beta\hbar\omega}.$$

$$\ln \left(\frac{\langle n \rangle + 1}{\langle n \rangle} \right) = \beta\hbar\omega.$$

Using these we have

$$\frac{S}{k_B} = \ln(\langle n \rangle + 1) + \langle n \rangle \ln \left(\frac{\langle n \rangle + 1}{\langle n \rangle} \right) = (\langle n \rangle + 1) \ln(\langle n \rangle + 1) - \langle n \rangle \ln \langle n \rangle.$$

Note that this formula for the entropy appears on pset 10.

40 Pressure of photon gas

a) Show that

$$P = -\partial_V U|_S.$$

b) Use this to show that for a photon gas in a cavity of volume V ,

$$P = - \sum_k \left(\langle n_k \rangle + \frac{1}{2} \right) \hbar \frac{d\omega_k}{dV}.$$

k here is a sum over modes in the cavity.

Note that this relies on the fact that the occupation numbers determine the entropy; since we want to fix the entropy, they don't get hit by the derivative.

c) Show that

$$\frac{d\omega_k}{dV} = -\frac{\omega_k}{3V}.$$

Use the fact that

$$\omega_k = c|\vec{k}| = c\sqrt{(n_x\pi/L)^2 + \dots} = \frac{\pi c}{V^{1/3}}\sqrt{n_x^2 + \dots}$$

d) Conclude that

$$P = +\frac{1}{3} \sum_k \left(\langle n_k \rangle + \frac{1}{2} \right) \hbar\omega_k \frac{1}{V} = \frac{1}{3} \frac{U}{V} + P_0.$$

Here P_0 is an additional *zero-point pressure*, coming from the sum over vacuum energies of the mode oscillators, which seems to be infinite. It in fact depends on the size of the box, and this dependence leads to Casimir forces which are measurable. (The simplest realization of this with real walls requires the walls to be conducting, and depends on some details of the interactions between the radiation and the walls, but this effect would also arise for periodic boundary conditions where there are no walls.)

41 Adsorption: an application of chemical potential

Let's consider a simple model of the condensation of vapor on the walls of its container; such a process where particles get stuck on the boundaries of their container is called *adsorption*.

Consider a volume V of vapor, composed of N classical particles, which has a boundary of area A . Each particle can be in the bulk of the volume, where its energy is

$$\epsilon_{bulk} = \frac{\vec{p}^2}{2m} \equiv \frac{p_x^2 + p_y^2 + p_z^2}{2m}.$$

Alternatively, if they want, they can stick to the walls. The word 'stick' is implemented by an energy gain of ϵ_0 when they are on the wall; however, on the wall they can only move in two of the three dimensions. For the part of the wall whose normal is the z direction, the 1-particle Hamiltonian is:

$$\epsilon_{wall} = \frac{p_x^2 + p_y^2}{2m} - \epsilon_0.$$

There is a similar expression for the other walls; we will not worry about corners, *i.e.* we'll take the walls of the container to be a big featureless plane. Note that we

suppose that they can still move around freely on the wall; this will not be such a great approximation for all examples of walls and gases.

Q: if we hold the whole container in thermal equilibrium at temperature T , what's the vapor pressure? If there were no opportunity to stick to the walls, this would be just the ideal gas pressure,

$$P = \frac{N}{V} k_B T.$$

In the problem at hand, the pressure should be decreased, since the total number of particles on the wall N_w plus the number in the bulk N_b is fixed to sum to N . The vapor is still an ideal gas but with fewer particles, so

$$P = \frac{N_w}{V} k_B T$$

where N_w will depend on N, V, A, ϵ_0 , and it is our job here to determine it.

Equilibrium occurs when the chemical potential for the particles on the walls equals that of the particles in the bulk:

$$\mu_{wall} = \mu_{bulk}.$$

Why is this? Well, the canonical partition sum is

$$\begin{aligned} Z &= \sum_{N_w=0}^N Z_{wall}(N_w) Z_{bulk}(N - N_w) \\ &= \sum_{N_w} e^{-\beta(F_{wall}(N_w) + F_{bulk}(N - N_w))} \\ &\approx e^{-\beta(F_{wall}(\langle N_w \rangle) + F_{bulk}(N - \langle N_w \rangle))} \end{aligned} \quad (41.1)$$

where in the last step we did the sum over N_w by saddle point, and $\langle N_w \rangle$ is determined by the saddle point condition

$$\begin{aligned} 0 &= \partial_{N_w} |_{T,V} (F_{wall}(N_w) + F_{bulk}(N - N_w)) |_{N_w = \langle N_w \rangle} \\ &= \partial_{N_w} |_{T,V} F_{wall}(N_w) - \partial_{N_b} |_{T,V} F_{bulk}(N_b) |_{N_w = \langle N_w \rangle, N_b = N - N_w} \end{aligned} .$$

See my notes from section on 04.07.2011 for a discussion of the saddle-point method. The large number which justifies its use here is the total number of particles.

So we'll need to evaluate the chemical potentials for each of the constituent systems. For the vapor, this is just the chemical potential for an ideal gas in three dimensions, which we did in lecture:

$$\mu_{vapor} = \partial_{N_b} |_{V,T} F_{bulk} = \partial_{N_b} |_{V,T} \left(-k_B T \frac{Z_{1,bulk}^{N_b}}{N_b!} \right)$$

$$= -k_B T \ln \frac{V}{N_b \lambda_{th}^3} \quad (41.2)$$

For the stuff on the walls, the partition function is $Z_{wall} = \frac{Z_{1,wall}^{N_w}}{N_w!}$ where the 1-particle partition function on the wall is

$$Z_{1,wall} = \int \frac{d^2 p d^2 x}{h^2} e^{-\frac{\beta}{2m}(p_x^2 + p_y^2) + \beta \epsilon_0} = A \left(\sqrt{\frac{2\pi m}{\beta h^2}} \right)^2 e^{\beta \epsilon_0} = \frac{A}{\lambda_{th}^2} e^{\beta \epsilon_0}.$$

So the chemical potential for the wall stuff is

$$\mu_{wall} = \partial_{N_w} F_{wall}|_{A,T} = \partial_{N_w}|_{A,T} \left(-k_B T \ln \left(\frac{Z_{1,wall}^{N_w}}{N_w!} \right) \right) \approx -k_B T \ln \frac{A}{N_w \lambda_{th}^2} - \epsilon_0.$$

Equating these two chemical potentials and exponentiating, we get

$$\frac{A}{N_w \lambda_{th}^2} e^{\beta \epsilon_0} = \frac{V}{N_b \lambda_{th}^3}$$

or

$$\frac{N_b}{V} = \frac{N_w}{A \lambda_{th}} e^{-\beta \epsilon_0}.$$

Now we must input the fact that the total number of particles is N , so this is an equation for N_b

$$\frac{N_b}{V} = \frac{N - N_b}{A \lambda_{th}} e^{-\beta \epsilon_0}$$

whose solution is a bit ugly

$$N_b = \frac{V N e^{-\beta \epsilon_0}}{A \lambda_{th} + V}.$$

So the vapor pressure is

$$P = \frac{N_b}{V} k_B T.$$

42 Another chemical potential example: liquid-vapor equilibrium

Say that the volume of a liquid made of N_{liq} particles is $V_{liq} = v_0 N_{liq}$; v_0 is the volume occupied by each particle in the liquid phase. Assume also that liquid particles enjoy some binding energy ϵ_0 that is absent if they are in the gaseous state. Compute the vapor pressure of a liquid in equilibrium with its own vapor, which we assume to be an ideal gas.

Set the chemical potentials equal: $\mu_{liq} = \mu_{vapor}$.

The answer I get is

$$P_{vapor} = \frac{N_{gas}}{V_{gas}} k_B T = \frac{k_B T}{v_0} e^{-\frac{\epsilon}{k_B T}}.$$

If V_{liq} is significant compared to V_{gas} we should include the fact that the volume available for the gas decreases, *i.e.* $V_{gas} = V - V_{liq}$ which will affect the answer for N_{gas} and hence the pressure.

43 Air

43.1 Law of isothermal atmospheres from canonical ensemble

Here let's give yet another perspective on the density of air as a function of height. Assume that the air is an ideal gas, with constant temperature T and assume that the gravitational potential experienced by one air molecule is $U(x, y, z) = mgz$, *i.e.* we are not so high that the curvature of the Earth matters, and m is the mass of an air molecule. We'd like to find the number density of air $n(z) = N/V$ as a function of height directly from the canonical ensemble.

Consider

$$p(z)dz = \text{prob}(\text{a molecule has height} \in (z, z + dz)).$$

Claim:

$$\frac{p(z)}{p(0)} = \frac{n(z)}{n(0)}.$$

The fact that the gas is assumed to be ideal means that we can compute the probability for one atom at a time. For a single atom we have

$$p(z)dz = \frac{dz \int dx dy dp_x dp_y dp_z e^{-\beta\epsilon(\vec{p}, z)}}{\int dz dx dy dp_x dp_y dp_z e^{-\beta\epsilon(\vec{p}, z)}}$$

where $\epsilon = \frac{\vec{p}^2}{2m} + mgz$. Most of the stuff cancels:

$$p(z)dz = \frac{e^{-\beta mgz} dz}{\int_0^\infty dz e^{-\beta mgz}} = \frac{e^{-\beta mgz} dz}{\zeta} ;$$

The denominator ζ is a constant which will cancel in ratios. So

$$\frac{p(z)}{p(0)} = \frac{e^{-\beta mgz} / \zeta}{e^{-\beta mg0} / \zeta} = e^{-\beta mgz} \implies n(z) = n(0) e^{-\beta mgz}.$$

43.2 Sound in air and adiabatic density variations

[Blundells §3.1]

The raging controversy of whether the physics of air is adiabatic versus isothermal continues in the study of sound propagation.

I claim that the velocity of sound in air is

$$v_s = \sqrt{\frac{B}{\rho}}$$

where $\rho = nm$ is the mass density, representing the inertia of a chunk of air, and

$$B \equiv -V \frac{\partial P}{\partial V} \quad (43.1)$$

is a compressibility representing some kind of restoring force to squishing the air. Note that since a pressure increase will decrease the volume, this quantity B is positive so v_s is real. Note that it B is also

$$B = \rho \frac{\partial P}{\partial \rho}. \quad (43.2)$$

This formula for the speed of sound I'll derive below. But what should we hold fixed when we take the derivative in (43.1)? Holding T fixed is what we would do if we thought that the temperature of the air was able to equilibrate rapidly on the timescale of the oscillations of the density involved in the soundwave. If heat doesn't get conducted that quickly then we should hold the entropy fixed. We can talk more about what's fast enough later.

Here's the derivation of the speed of sound. It can be derived from

(1) the continuity equation, *i.e.* the statement that mass can't go away, so if the mass density changes somewhere it's because some mass left or arrived

$$-\partial_t \rho = \vec{\nabla} \cdot (\rho \vec{u}).$$

(2) Newton's law, which in this context is called the Euler equation, and says that a pressure gradient results in an acceleration:

$$-\vec{\nabla} p = \rho \frac{D\vec{u}}{Dt} \equiv \partial_t \vec{u} + (\vec{u} \cdot \vec{\nabla}) \vec{u}.$$

The funny thing on the right is called a 'convective derivative' and I'm not going to explain it here. The funny second term will be negligible for low-amplitude density variations. We consider small ripples about an equilibrium configuration where the density of air is ρ and its velocity is zero.

And let's do it in one dimension where things are simple. The continuity says

$$-\partial_t \rho = \partial_x(\rho u) = u \partial_x \rho + \rho \partial_x u$$

Define $s \equiv \frac{\delta \rho}{\rho}$ which is the small fractional deviation of the density. Then we have

$$u \partial_x s + \partial_x u = -\partial_t s.$$

The first term on the LHS is a product of two small things and so we ignore it. We have

$$\partial_x u = -\partial_t s. \tag{43.3}$$

Euler's equation then says

$$\partial_t u = -\frac{1}{\rho} \partial_x p = -\frac{B}{\rho} \partial_x s$$

where we used (43.2).

Finally, eliminate u from these equations using (43.3):

$$\partial_t(\partial_x u) = \partial_x \left(-\frac{B}{\rho} \partial_x s \right)$$

which says, again ignoring terms quadratic in small things,

$$\partial_t^2 s = \frac{B}{\rho} \partial_x^2 s$$

which is the wave equation with the advertised speed. (If you don't believe me, plug in $s \propto e^{i(kx - \omega t)}$ and see what relation you get between ω and k .)

44 Pressure of quantum gases

a) Show that

$$P = -\partial_V U|_S.$$

This follows from the first law $dU = TdS - PdV$.

b) Use this to show that for an ideal quantum gas (*e.g.* thermal radiation) in a cavity of volume V ,

$$P = - \sum_k \left(\langle n_k \rangle + \frac{1}{2} \right) \hbar \frac{d\omega_k}{dV}. \tag{44.1}$$

k here is a sum over modes in the cavity. Here $\langle n_k \rangle$ is the occupation number of the mode with wavenumber k ; its for is determined by whether the particles in the gas are fermions

$$\langle n_k \rangle = \frac{1}{e^{\beta(\epsilon(k) - \mu)} + 1}$$

or bosons

$$\langle n_k \rangle = \frac{1}{e^{\beta(\epsilon(k)-\mu)} - 1}$$

or maxwell-boltzmons.

$$\langle n_k \rangle = e^{-\beta(\epsilon(k)-\mu)}.$$

Note that the dependence on k arises only through the dependence on the single-particle energy $\epsilon(k)$.

The result (44.1) relies on the fact that the occupation numbers determine the entropy; we showed this last week for a single harmonic oscillator, but this is in fact a general statement: the occupation numbers specify the state. Since we want to fix the entropy, they don't get hit by the derivative⁷.

c) Show that for ultra-relativistic particles

$$\frac{d\omega_k}{dV} = -\frac{\omega_k}{3V}.$$

Use the fact that

$$\omega_k = c|\vec{k}| = c\sqrt{(n_x\pi/L)^2 + \dots} = \frac{\pi c}{V^{1/3}}\sqrt{n_x^2 + \dots}$$

d) Conclude that

$$P = +\frac{1}{3} \sum_k \left(\langle n_k \rangle + \frac{1}{2} \right) \hbar\omega_k \frac{1}{V} = \frac{1}{3} \frac{U}{V} + P_0.$$

Here P_0 is an additional *zero-point pressure*, coming from the sum over vacuum energies of the mode oscillators, which seems to be infinite. It in fact depends on the size of the box, and this dependence leads to Casimir forces which are measurable. (The simplest realization of this with real walls requires the walls to be conducting, and depends on some details of the interactions between the radiation and the walls, but this effect would also arise for periodic boundary conditions where there are no walls. Note that we must subtract off an infinite vacuum energy *density* to get a finite answer for the pressure.)

e) Show that for quantum gas of non-relativistic particles ($\epsilon_k = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$),

$$P = \frac{2}{3} \frac{U}{V}.$$

⁷ For a better argument see the notes from May 12, 2011, page 1. (§47)

45 Johnson-Nyquist noise of an electrical circuit at finite temperature

Consider a driven electrical circuit which is at temperature T – it is difficult to prevent circuit elements from getting hot. Thermal fluctuations will cause the current through the circuit to vary from the value one would calculate using Ohm’s law. Can we quantify this?

To be a little more precise, let’s suppose we have two conductors, each of resistance R connected in series, with no battery or anything. The thermal motion of the charges in one resistor will cause a current to flow in the other, which is $I = \frac{V}{2R}$, where V is the EMF created by the thermal motion.

Here is a trick (due to Nyquist, whose 1928 paper is posted on Dave Litster’s site) which makes finding the fluctuations of the EMF $\langle V^2 \rangle$ into an application of thermal radiation. We take a detour to discuss blackbody radiation in one dimension.

45.1 Thermal radiation in one dimension

Consider EM waves confined to propagate in only one dimension, as in a transmission line or waveguide (think coaxial cylinders) of length L . There are still two polarizations.

Assume that the boundary conditions at the two ends demand that the amplitude vanish at $x = 0, L$. (This is the case if the circuit elements to which the transmission line is attached are *impedance matched*.) Then the allowed wavenumbers are $k = \frac{n\pi}{L}, n = 1, 2, \dots$. Therefore

$$D(k)dk = \frac{\text{number of dots}}{\text{volume in k-space occupied by a dot}} = \frac{dk}{L/\pi} = \frac{L}{\pi} dk.$$

As usual for photons, $\omega = ck$, so the density of states in frequency is determined by

$$D(k)dk = \frac{Ld\omega}{\pi c} = D(\omega)d\omega .$$

The energy per unit length in the transmission line in thermal equilibrium is

$$\frac{U}{L} = \frac{1}{L} \int_0^\infty d\omega D(\omega) \langle \epsilon(\omega) \rangle = \frac{1}{L} \int_0^\infty \frac{L}{\pi c} \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} .$$

The energy per unit length in frequencies between ω and $\omega + d\omega$ is then

$$\frac{dU(\omega)}{L} = \frac{\hbar\omega}{\pi c} \frac{1}{e^{\beta\hbar\omega} - 1} .$$

Now these standing waves in the transmission line can also be considered as a bunch of radiation (half of it) traveling to the right at the speed of light superposed with the same amount of radiation traveling to the left at the speed of light. The energy incident on the $x = L$ end of the line per unit time (in the given frequency range) is then

$$\frac{c}{2} \frac{dU(\omega)}{L} = \frac{\hbar\omega}{2\pi} \frac{1}{e^{\beta\hbar\omega} - 1} \stackrel{\beta\hbar\omega \ll 1}{\approx} \frac{k_B T}{2\pi} d\omega$$

where the last step, where we take the high-temperature limit, reproduces classical equipartition.

45.2 Back to circuit noise

To use what we've just learned, consider connecting an *impedance-matched* transmission line between the two resistors. I don't want to get into complications about waveguides here : what this means is that the electromagnetic field has Dirichlet boundary conditions at the end of the transmission line, as we assumed above.

If this energy is incident on a resistor with resistance R in equilibrium at temperature T , the power going in and being dissipated is

$$\langle I^2 R \rangle.$$

To relate this to the voltage fluctuations, we need to relate the current to the voltage, which we did above, $I = V/2R$, so

$$\langle I^2 R \rangle = \left\langle \frac{V^2}{4R} \right\rangle = P.$$

We can refine this answer by asking about how much current and voltage there is in each frequency band⁸ – the noise voltage is like an AC voltage source. Then we have

$$P(\nu) d\nu = \left\langle \frac{V^2}{4R} \right\rangle d\nu$$

(with $\nu \equiv \omega/2\pi$) and therefore

$$\langle V^2 \rangle_\omega = R \frac{2}{\pi} \frac{\hbar\omega d\omega}{e^{\beta\hbar\omega} - 1} \stackrel{\beta\hbar\omega \ll 1}{\approx} 4Rk_B T d\nu.$$

This is an example of a very general result called the *fluctuation-dissipation theorem* which relates the size of fluctuations in thermal equilibrium ($\langle V^2 \rangle$) to the amount of dissipation (R).

⁸ We can argue that the condition of thermal equilibrium between the resistors and the radiation must be met independently for each frequency as follows. If it weren't the case, and some frequencies had more radiation going one way than the other you could put a filter (some inductors and capacitors) which would block those frequencies and make heat flow between the resistors; since they are at the same temperature by assumption this would violate the second law.

46 Warmup question: entropy and occupation numbers

Consider a quantum ideal gas of bosons or fermions in a weird box that has just *a single mode* of energy ϵ . Don't fix the number of particles. What is the entropy?

In a general, non-weird, box the partition function for a quantum ideal gas is

$$Z = \sum_{\text{states}} e^{-\beta E_{\text{state}}} = \prod_{\text{modes, } k} \sum_{n_k} e^{-\beta \sum_k n_k \epsilon_k}.$$

Here I've used the fact that we can enumerate the states of a collection of quantum identical particles by giving the occupation number of each single-particle mode. (I also subtracted off the vacuum energy $\sum_k \frac{1}{2} \epsilon_k$ for convenience.)

The sum over n_k runs over $0, 1, 2, \dots, \infty$ for bosons, and over $0, 1$ for fermions.

For bosons in a cavity with one mode, the partition sum is

$$Z_B = \sum_{n=0}^{\infty} e^{-\beta \epsilon n} = \frac{1}{1 - e^{-\beta \epsilon}}$$

– it's just a single simple harmonic oscillator. I discussed the entropy of an SHO in the notes previously (on 04.28.2011).

For fermions in the weird cavity,

$$Z_F = \sum_{n=0}^1 e^{-\beta \epsilon n} = 1 + e^{-\beta \epsilon}.$$

This just a two-state system. The average occupation number is

$$\langle n \rangle_F = \frac{\sum_{n=0}^1 n e^{-\beta \epsilon n}}{Z_F} = \frac{1}{e^{\beta \epsilon} + 1}$$

the Fermi-Dirac distribution with $\mu = 0$. This function satisfies some identities analogous to those for the bose factor:

$$1 - \langle n \rangle_F = \frac{1}{1 + e^{-\beta \epsilon}} = \frac{1}{Z_F}.$$

$$\ln \frac{1 - \langle n \rangle_F}{\langle n \rangle_F} = \beta \epsilon.$$

These identities are useful in showing that the entropy of the one-mode Fermi gas is

$$\frac{s_F}{k_B} = -((1 - \langle n \rangle_F) \ln(1 - \langle n \rangle_F) + \langle n \rangle_F \ln \langle n \rangle_F).$$

Notice that this is of the form

$$s = - \sum_{\alpha} p_{\alpha} \ln p_{\alpha}$$

where α is a sum over possibilities (fermion present or absent) with probability $p = \langle n \rangle_F$ and $1 - p$ respectively. This is just the Shannon entropy of the probability distribution.

47 Comment about pressure of ideal gases

In a previous recitation I gave an argument that a non-relativistic ideal gas has a pressure that satisfies

$$P = \frac{2}{3} \frac{E}{V}$$

(and an ultrarelativistic ideal gas has $P = \frac{1}{3} \frac{E}{V}$). A better argument than the one I gave appears in Baierlein (on page 97).

The basic fact which is relevant is that

$$P = \sum_J \left(\frac{-\partial E_J}{\partial V} \right) \frac{e^{-\beta E_J}}{Z}, \quad (47.1)$$

where J is a sum over states of the many-body system – i.e. it is a multi-index which runs over sets of single-particle-state occupation numbers $J = \{n_{\alpha}\}$ – and

$$E_{\{n_{\alpha}\}} = \sum_{\alpha} \epsilon_{\alpha} n_{\alpha}.$$

From this formula, the analysis is just as we discussed before.

The formula (47.1) is true by the following nice argument.

Consider the system in a state J . Expand the volume infinitesimally by dV , while fixing the state (*i.e.* the occupation numbers) – in the quantum context, this is what it means to do something adiabatically. The system exerts a pressure on its walls and does work. The energy of the state $E_J(V)$ will decrease. Energy conservation says

$$0 = dE = (\text{pressure in state } J) dV + \frac{\partial E_J}{\partial V} dV.$$

Now average the pressure in state J using the canonical ensemble to get (47.1).

48 Bose condensation in two dimensions?

Can Bose condensation happen in two dimensions, at nonzero temperature, in the thermodynamic limit?

This would mean that

$$N = \langle n_1 \rangle + \sum_{\alpha \geq 2} \langle n_\alpha \rangle$$

with a $\langle n_1 \rangle = N/b$ with b a constant which stays finite as $N \rightarrow \infty$. Here

$$\langle n_\alpha \rangle = \frac{1}{e^{\beta(\epsilon_\alpha - \mu)} - 1}$$

is the Bose distribution.

Let's focus on non-relativistic spin-0 bosons. For convenience let's fix the zero of energy so that the groundstate has $\epsilon_1 = 0$. Then

$$\langle n_1 \rangle = \frac{1}{e^{-\beta\mu} - 1} = \frac{z}{1 - z}, \quad z \equiv e^{\beta\mu};$$

z is sometimes called the *fugacity*. BEC would mean

$$\langle n_1 \rangle = \frac{z}{1 - z} = N/b \implies z = \frac{1}{1 + b/N} \approx 1 - b/N,$$

where in the last step we used the fact that b is some number like 2 and N is like a billion, *i.e.* going to infinity in the thermodynamic limit.

Our accounting of all the particles then says

$$N = \frac{z}{1 - z} + \int_{0+}^{\infty} \frac{D_2(\epsilon) d\epsilon}{e^{\beta(\epsilon - \mu)} - 1}.$$

In two dimensions, we have

$$D_2(\epsilon) d\epsilon = D_2(k) dk = \frac{\frac{1}{4} 2\pi k dk}{(\pi/L)^2} = \frac{L^2 m}{2\pi \hbar^2} d\epsilon.$$

In the last step we used the NR dispersion relation

$$\epsilon = \frac{\hbar^2 k^2}{2m} \implies d\epsilon = \frac{\hbar^2}{m} k dk.$$

Notice that the density of states in energy is *constant*.

So

$$N = \frac{z}{1 - z} + \frac{L^2 m}{2\pi \hbar^2} \int_{0+}^{\infty} \frac{d\epsilon}{z^{-1} e^{\beta\epsilon} - 1}$$

$$\begin{aligned}
&= \frac{z}{1-z} + \frac{L^2 m}{2\pi\hbar^2\beta} \int_{0^+}^{\infty} \frac{dx}{z^{-1}e^x - 1} \\
&\equiv \frac{z}{1-z} + \frac{L^2 m}{2\pi\hbar^2\beta} \mathcal{I}_2(z).
\end{aligned} \tag{48.1}$$

$$\mathcal{I}_2(z) = \int_{0^+}^{\infty} \frac{dx}{z^{-1}e^x - 1} = \int_{0^+}^{\infty} \frac{dx z e^{-x}}{1 - z e^{-x}} = \int_{0^+}^{\infty} \sum_{l=1}^{\infty} (z e^{-x})^l$$

where we used $\frac{a}{1-a} = \sum_{l=1}^{\infty} a^l$

$$\mathcal{I}_2(z) = \sum_{l=1}^{\infty} z^l \int_{0^+}^{\infty} e^{-lx} = \sum_{l=1}^{\infty} \frac{z^l}{l} = -\ln(1-z).$$

Dividing by N , the accounting is now

$$\begin{aligned}
1 &= \frac{1}{b} + \frac{m}{2\pi\hbar^2\beta} \frac{L^2}{N} \mathcal{I}_2(z) \\
&= \frac{1}{b} - \frac{m}{2\pi\hbar^2} k_B T \frac{L^2}{N} \ln(1-z) \\
&= \frac{1}{b} + \frac{m}{2\pi\hbar^2} k_B T \frac{L^2}{N} \ln(N/b)
\end{aligned} \tag{48.2}$$

But the second term on the RHS goes to infinity when we take N large while fixing the density N/L^2 . The excited states can accommodate an arbitrarily large number of the particles.

Notice that if we take $T = 0$ from the beginning indeed all the particles sit in the groundstate, $\langle n_1 \rangle = N, b = 1$.

The fact that there is no Bose condensation in 2d and the fact that there are no solids in 2d are not completely unrelated. In both cases the failure is by some logarithmic (*i.e.* very slowly growing) divergence, so each of these things are in fact approximately possible in practice. But more fundamentally, each of them involves correlations over very long distances. Despite our familiarity with them, a solid is actually a very shockingly weird thing: if you know the position of an atom at one end, you can make a pretty good prediction for the position of an atom all the way at the other end of the object. Such a state is said to have *long range order*. Similarly, in a Bose condensate, all of the particles have the same wave function and so the shape of the wavefunction relates the behavior of things that are far apart. Two dimensions is not a hospitable environment for long range order because of the growth of fluctuations.

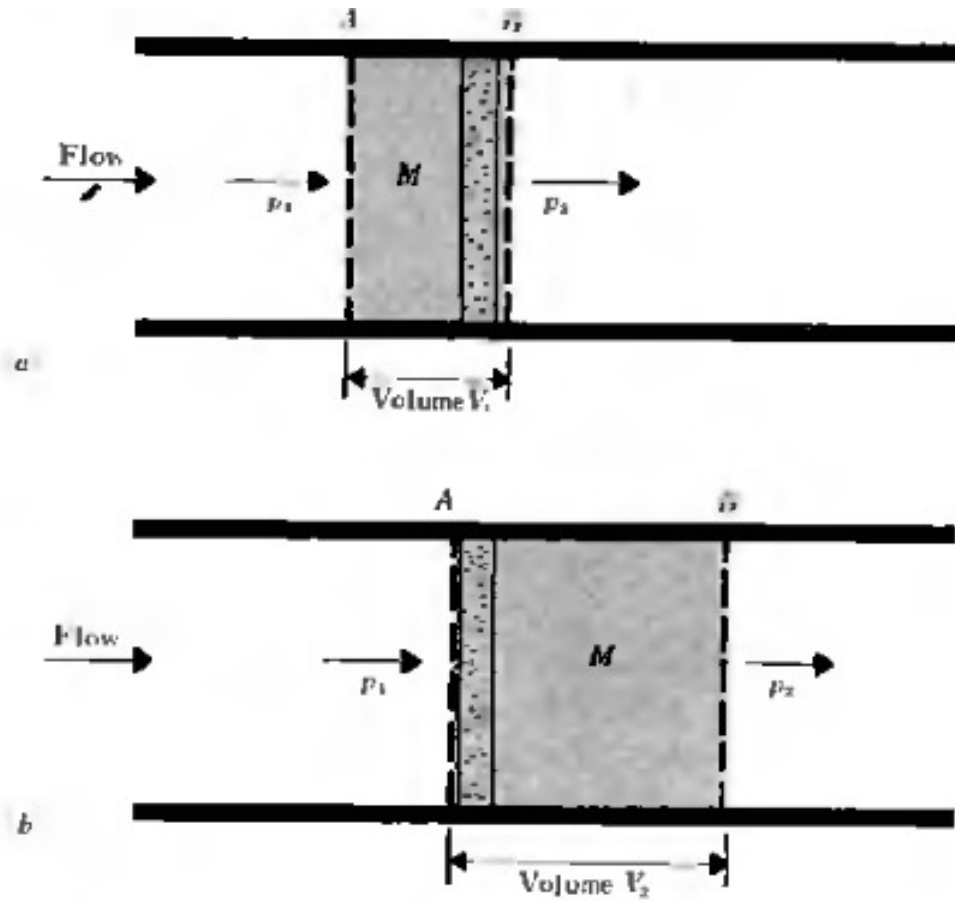


Fig. 5·10·2 Diagram showing a mass M of gas passing through a constriction (a porous plug in this case) (a) before passing through the constriction (b) after passing through the constriction.

Figure 12: From Reif.

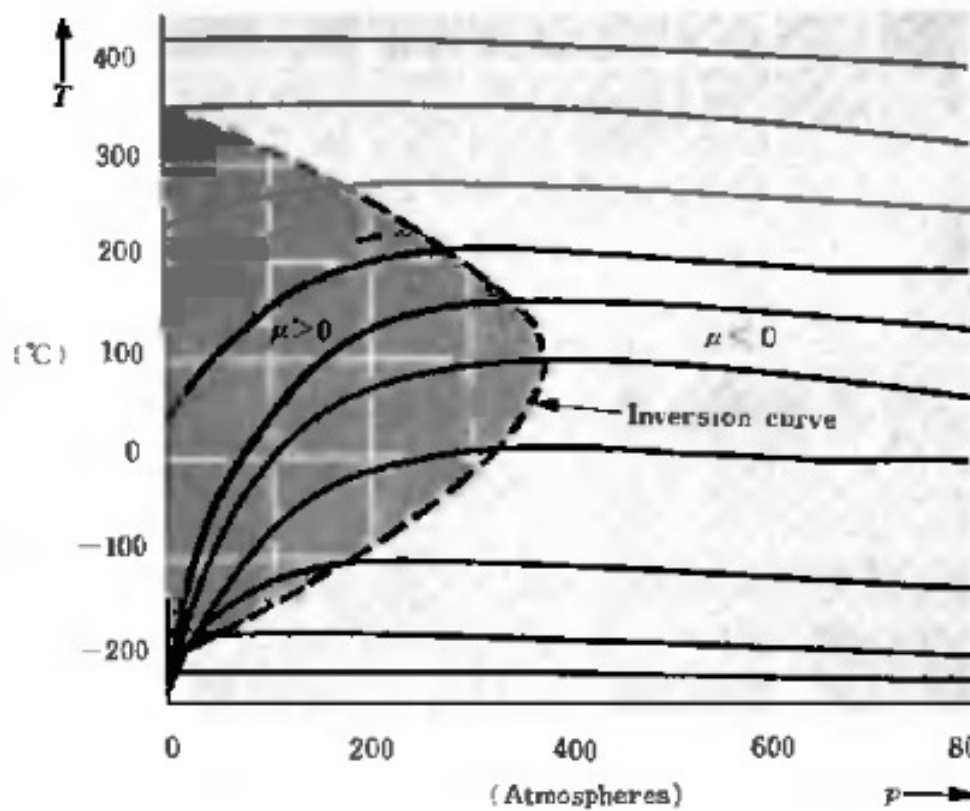


Fig. 5·10·3 Curves of constant enthalpy H in the pT plane of a gas. The numerical values are for nitrogen (N_2). The dashed line is the inversion curve.

Figure 13: From Reif.

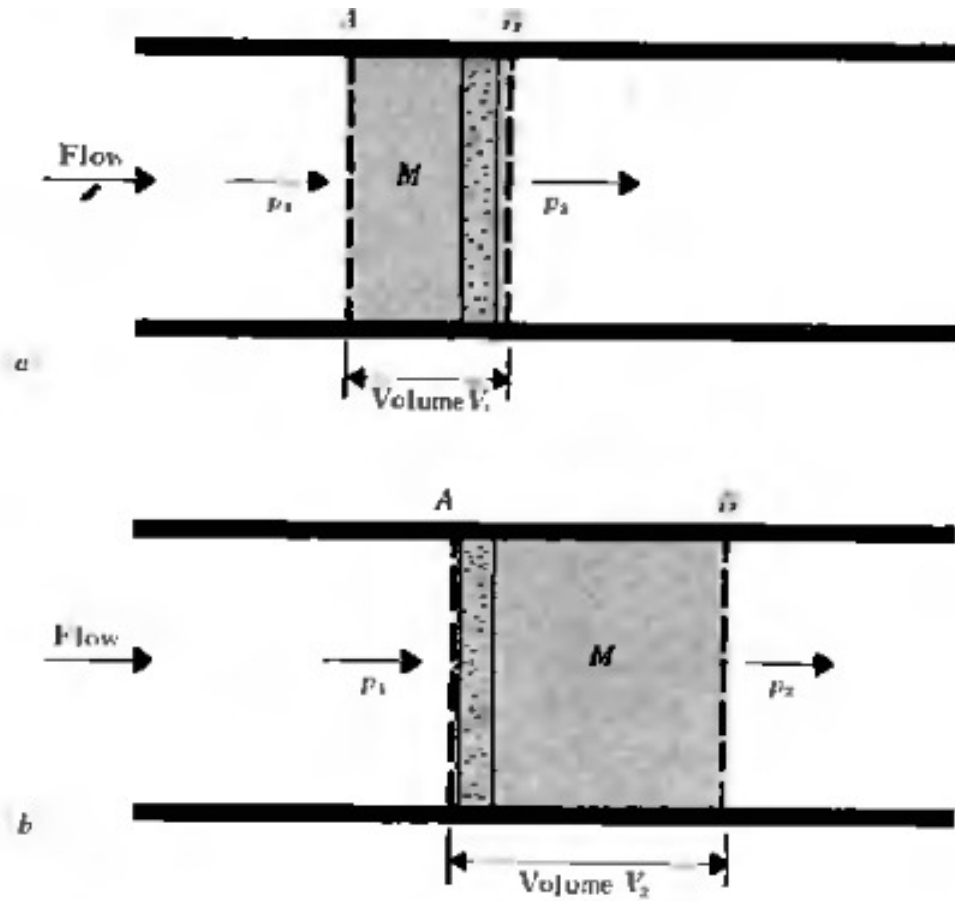


Fig. 5·10·2 Diagram showing a mass M of gas passing through a constriction (a porous plug in this case) (a) before passing through the constriction (b) after passing through the constriction.

Figure 14: From Reif.

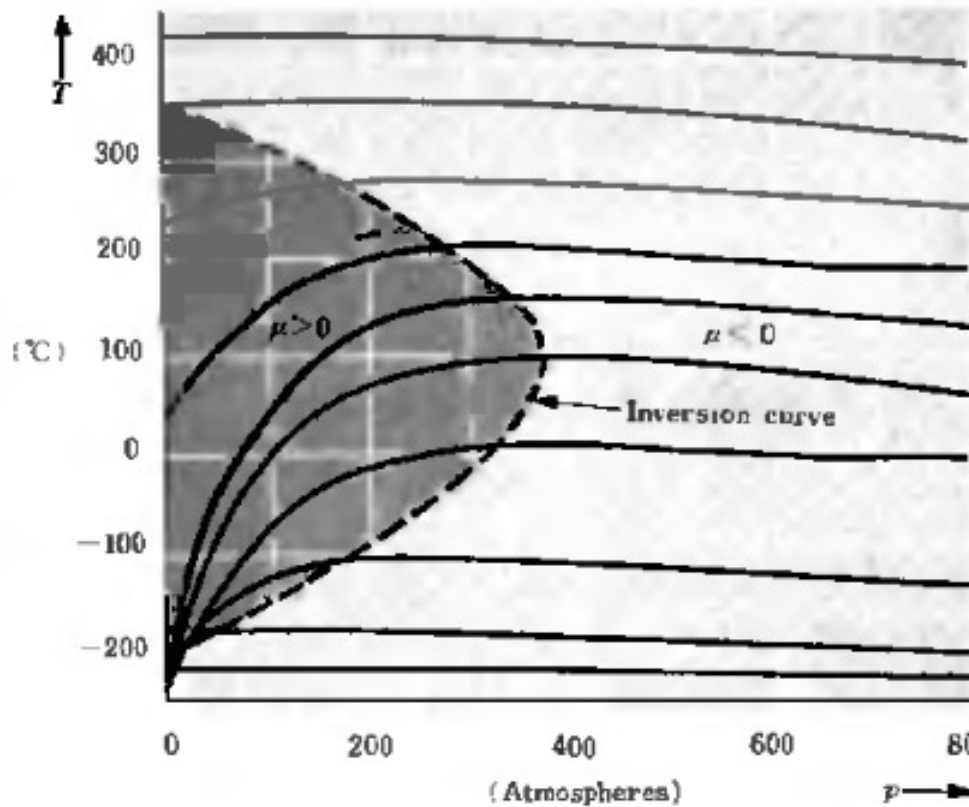


Fig. 5·10·3 Curves of constant enthalpy H in the pT plane of a gas. The numerical values are for nitrogen (N_2). The dashed line is the inversion curve.

Figure 15: From Reif.

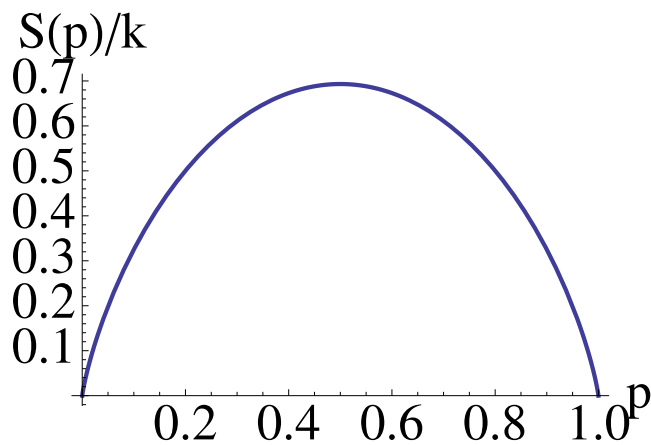


Figure 16: The Shannon entropy for the Bernoulli (two-outcome) process with probability p .

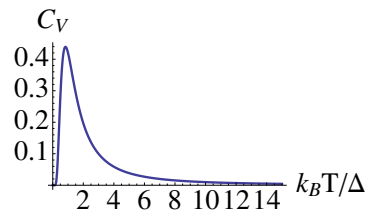


Figure 17: Specific heat of a two-state system as a function of temperature.

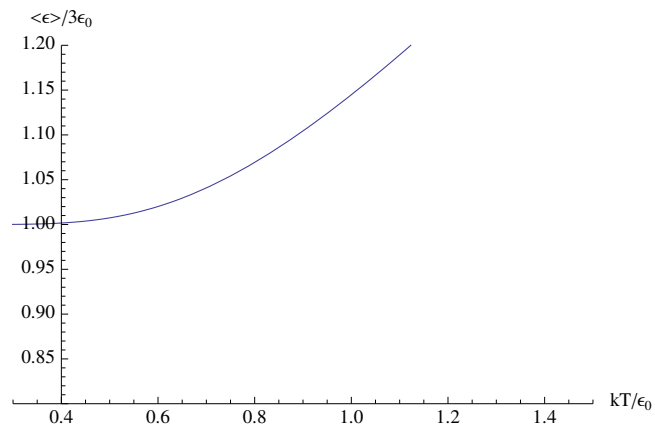


Figure 19: The energy interpolates between the quantum groundstate energy at $T = 0$ and the linear- T ideal gas behavior at high temperature.