

Chapter 3

The Method of the Most Probable Distribution

3.1 Averages and Perfect Classical Gases

In the introduction we noted that statistical mechanics has a single purpose: to find the distribution of N weakly interacting systems over the energy states available to them. That is, how many, n_s , of the N systems in the assembly do we expect to find in the state of s having energy ϵ_s . With this distribution we can calculate the average properties of the systems. For example, the average energy of the system is

$$\frac{\sum_s n_s \epsilon_s}{\sum_s n_s} = \bar{\epsilon} \quad (3.1)$$

We could also introduce the probability of finding a single system in state s , $f_s = n_s/N$ so that the average is

$$\sum_s f_s \epsilon_s = \bar{\epsilon} \quad , \quad \sum_s f_s = 1 \quad (3.2)$$

Finding the distribution of systems over their possible states s is really a mathematical problem. This mathematical problem has wide application and may be interpreted physically in different ways. For example, what are the systems and what is the assembly? To solve the mathematical problem we use the method of the most probable distribution, also referred to the Boltzmann combinatorial method, developed by Boltzmann (1847-1906) in 1877. There are other methods, such as the method of steepest descents developed by Darwin and Fowler which is more rigorous. But what the combinatorial method lacks in rigor is made up in simplicity leaving us free to focus on concepts.

Boltzmann was primarily interested in gases. He contributed perhaps more than anyone else to the understanding of them in the late 19th century. In a gas at STP, the average interatomic (or molecular) spacing is approximately

twenty times the atomic dimensions. Thus the atoms are essentially independent and collide only rarely with one another. They are weakly interacting systems. This is strictly true only for an interatomic interaction of short range and may not apply, for example, to a gas of electrons interacting via the long range coulomb interaction. At present we exclude these cases. However, some interaction is needed so that energy can be exchanged between the systems maintaining them in weak "thermal" contact. In Boltzmann's application of his combinational method to gases, the atom is the system and the assembly of systems is the gas itself. Thus to give physical content to the mathematical problem we can have in mind a gas (assembly) of N weakly interacting atoms (systems) and seek the distribution of the atoms over their possible energy states.

A perfect or ideal gas is a gas satisfying the above conditions of weak interaction between atoms for which the interatomic potential energy of interaction is assumed to be exactly zero. We therefore consider a classical, perfect gas. We also suppose the gas is isolated (U is constant and the number of atoms N is fixed). Since the system is an atom, the system phase space has only 6 dimensions (for a monatomic gas). For a polyatomic gas, for which each 'molecule' possesses f degrees of freedom (3 translational and $f-3$ internal), the system phase space is $6f$ dimensional, but for simplicity we consider only monatomic gases at present. From section 1.4 an element of phase space is then

$$d\Gamma = dp_x dp_y dp_z \cdot dx dy dz \quad (3.3)$$

We may specify the state of an atom system by specifying its momentum \vec{p} and position \vec{r} in phase space. We could then seek the distribution of the atoms over the possible cells or regions in phase space to find the most probable distribution. However, it is more general, and simpler, to evaluate the distribution over the possible energy states ϵ_s of the atoms. For the special case of the classical perfect gas the two distributions amount to the same thing since momentum and energy are directly connected by $\epsilon = p^2/2m$ and, since the potential energy is zero, no position in space is preferred.

In the Gibbs interpretation of our mathematical model, the system is the whole gas, liquid or solid consisting of M particles (bodies, atoms) that is under consideration. The ensemble of systems is N mental copies of the M body system. The states s are states of the whole M body system and the ϵ_s are the corresponding energy states of the whole M body system. The copies are assumed to be weakly interacting and to make up the heat bath at temperature T for the system under consideration. As discussed at the end of section 1.2, the state of the system can also be specified by specifying the region,

$$d\Gamma = \prod_{i=1}^M d\vec{p}_i d\vec{x}_i.$$

that the system occupies in $6M$ dimensional phase space, Γ .

3.2 The Combinatorial Method

To find the distribution of the systems over their possible energy states, we enumerate the states, $\epsilon_1, \epsilon_2, \dots, \epsilon_s \dots$ beginning with the lowest state first and ending with the highest state ϵ_r . We then arbitrarily assign n_1 systems to state ϵ_1, n_2 to ϵ_2 and so on, i.e.

State No.	1	2	3	...	s	...	r	
Energy	ϵ_1	ϵ_2	ϵ_3	...	ϵ_s	...	ϵ_r	(3.4)
Number	n_1	n_2	n_3	...	n_s	...	n_r	

When applying this enumeration to a classical gas where ϵ is continuous we must divide the energy up into cells. The cell size we chose was 1 state for each cell of dimensions $\Delta \vec{p} \Delta \vec{x} = h^3$. For a gas at STP this cell size is so small that, on average, there is roughly 1 chance in 100,000 that a cell is occupied. Hence n_s is zero mostly and occasionally one or two. However, in Boltzmann's combinatorial method the n_s must be large since it is assumed the n_s changes smoothly from state to state and differentials are employed. This led many to reject Boltzmann's method when he first introduced it. However, it is possible to combine enough cells to make each n_s large and successfully repeat the derivation (see, for example, Sommerfeld). This difficulty does not arise in Gibb's interpretation (Since N can be infinite) and we will not get embroiled in this problem here. To complete the enumeration we must add the restrictions

$$\sum_{s=1}^r n_s = N, \quad \sum_{s=1}^r n_s \epsilon_s = U \tag{3.5}$$

What is the probability of observing a distribution with say n_1 atoms in ϵ_1, n_2 in $\epsilon_2 \dots n_s$ in ϵ_s and so on? Firstly, we use the assumption of equal a priori probability of phase space (Section 1.5). This tells us that each state or cell is equally likely to be occupied. No state is intrinsically preferred over another. Then, following Boltzmann we assume the probability of seeing the distribution is simply proportional to the number of ways the distribution can be formed. This is, the number of permutations of the particles among the states which leads to the same distribution is proportional to the probability of observing this distribution. In other words the permutability is a measure of the probability. The number of ways the distribution enumerated in 3.4 can be formed is¹

$$W = \frac{N!}{n_1! n_2! n_3! \dots n_r!} \tag{3.6}$$

¹Note that this expression assumes that the atoms are distinguishable. If the particles are not distinguishable, then we should divide by $N!$, since permutation does not lead to a new state. This change would not affect the final result.

Here $N!$ is the total number of permutations of the atoms. Then we must divide by the number of ways we can permute the n_s within state s since permuting the systems within a state does not lead to a new distribution of the systems over the states.

To maximize W with respect to n_s , we need a convenient analytic expression for $N!$ and $n!$. This is provided by Stirling's approximation

$$N! = \left(\frac{N}{e}\right)^N \quad (3.7)$$

A derivation of Stirling's approximation is

$$\begin{aligned} \log N! &= \log 1 + \log 2 + \dots + \log N = \int_1^N \log x \, dx \\ &= x(\log x - 1) \Big|_1^N = N(\log N - 1) - (\log 1 - 1) \\ &= N(\log N - 1) + 1 \\ &\approx N(\log N - 1) \end{aligned}$$

which is an excellent approximation if N (and n_s) is large. We choose to maximize $\log W$ rather than W ,

$$\begin{aligned} \log W &= \log N! - \sum_{s=1}^r \log n_s! \\ &= N(\log N - 1) - \sum_{s=1}^r n_s(\log n_s - 1) \end{aligned}$$

Although we must maximize $\log W$ keeping the number of systems N and the total energy U constant, it is an interesting exercise to first find the most probable distribution unrestricted by constant energy. For this we simply seek that W which is an extremum with respect to n_s . At the extremum $\delta \log W = 0$, i.e.

$$\begin{aligned} \delta \log W &= - \sum_{s=1}^r \delta n_s \log n_s - \sum_{s=1}^r \delta n_s + \sum_{s=1}^r \delta n_s \\ &= - \sum_{s=1}^r \delta n_s \log n_s = 0 \end{aligned}$$

Since N is constant, this variation is restricted so that

$$\delta N = \sum_s \delta n_s = 0. \quad (3.8)$$

The variations δn_s have to sum to zero. To satisfy 3.8 and make $\delta \log W$ vanish requires that the most probable distribution be that for which each $\log n_s$ is equal. This confirms our original hypothesis of equal a priori probability of phase space. That is, with no restrictions on the total energy, each cell (energy) in phase space is equally likely to be occupied.

To introduce formally the restrictions 3.5 of constant N and U , we add N and U to $\log W$ each multiplied by an arbitrary Lagrange multiplier (α and β). Then the most probable distribution is that for which

$$\delta (\log W - \alpha \sum_s n_s - \beta \sum_s \epsilon_s n_s) = 0$$

Since α and β are arbitrary, the δN and δU must be zero independently above which guarantees constant N and U . Varying with respect to the n_s gives

$$- \sum_s \delta n_s (\log n_s + \alpha + \beta \epsilon_s) = 0$$

Since each δn_s is arbitrary each term in the sum must vanish and the special values \bar{n}_s which do this are

$$\log \bar{n}_s + \alpha + \beta \epsilon_s = 0$$

or

$$\bar{n}_s = e^{-\alpha} e^{-\beta \epsilon_s} \quad (3.9)$$

Inserting these special values of n_s into $\log W$ gives the maximum value of W as

$$\begin{aligned} \log W_{max} &= N \log N - \sum_{s=1}^r \bar{n}_s \log \bar{n}_s \\ &= N \log N + \alpha \sum_{s=1}^r \bar{n}_s + \beta \sum_{s=1}^r \epsilon_s \bar{n}_s \end{aligned} \quad (3.10)$$

We now turn to identifying β and making contact with thermodynamics.

3.3 The Boltzmann Factor, Partition Function and Thermodynamics

Using the method of the most probable distribution, we found above that the expected number of systems in state s having energy ϵ_s is

$$\bar{n}_s = e^{-\alpha} e^{-\beta \epsilon_s}. \quad (3.9)$$

With $\beta = (kT)^{-1}$ this is an example of the famous Boltzmann factor, derived first by Boltzmann in (1868). The Boltzmann factor gives the probability $p(s)$ of finding a system in state s as $p_s \propto e^{-\epsilon_s/kT}$ or properly normalized,

$$p_s = \frac{e^{-\epsilon_s/kT}}{\sum_s e^{-\epsilon_s/kT}}.$$

The Boltzmann factor gives the probability for observing a system in energy state ϵ_s for both quantum and classical states. However, only in classical statistical mechanics is the expected number of particles in each state (the occupation number) also proportional to $e^{-\epsilon_s/kT}$.

At this point α and β are arbitrary Lagrange multipliers. We now interpret α and β by regarding the systems as point particles and the assembly as an isolated gas of N particles at constant energy U . This is the Boltzmann interpretation of the mathematical model. The Gibbs interpretation is made in Chapter 5 and the determination of α and β within that interpretation is made there. Those wishing to consider only the Gibbs interpretation applicable to all systems should skip the following and move on to Chapter 5.

For a classical gas of N particles at constant energy U , we have

$$N = \sum_s \bar{n}_s = e^{-\alpha} \sum_s e^{-\beta \epsilon_s}$$

$$U = \sum_s \bar{n}_s \epsilon_s = e^{-\alpha} \sum_s e^{-\beta \epsilon_s} \epsilon_s$$

We take the gas through thermodynamic changes of state and identify, in the changes, the relation between the microscopic distribution of the particles over the energy states and the thermodynamic functions. To begin, we define the partition function, Z_o , or "sum over states"

$$Z_o = \sum_s e^{-\beta \epsilon_s} \quad (3.11)$$

The Z stands for the German Zustandssumme. We emphasize that since the system here is a particle, Z_o is the single particle partition function and ϵ_s are the single particle (translational) energy states. We then eliminate α in favor of Z_o ,

$$e^{-\alpha} = \frac{N}{Z_o} \quad (3.12)$$

so that

$$\bar{n}_s = \frac{N}{Z_o} e^{-\beta \epsilon_s} = -\frac{N}{\beta} \frac{\partial}{\partial \epsilon_s} \log Z_o \quad (3.13)$$

$$U = \frac{N}{Z_o} \sum_s e^{-\beta \epsilon_s} \epsilon_s = -N \frac{\partial}{\partial \beta} \log Z_o$$

To identify β , we take the gas through a small change of state and watch the change in $N \log Z_o$. Since from (3.13) Z_o is a function of β and the ϵ_s , in a reversible change,

$$\begin{aligned}
d(N \log Z_o) &= N \frac{\partial}{\partial \beta} (\log Z_o) d\beta + N \sum_s \frac{\partial}{\partial \epsilon_s} (\log Z_o) d\epsilon_s \\
&= -U d\beta - \beta \sum_s \bar{n}_s d\epsilon_s \\
&= -d(\beta U) + \beta dU - \beta \sum_s \bar{n}_s d\epsilon_s
\end{aligned} \tag{3.14}$$

From Section 2.1, we recall that to change the energy states of the systems in the gas the work done by the gas is $dW = -\sum_s \bar{n}_s d\epsilon_s$. Hence

$$\frac{1}{\beta} d(N \log Z_o + \beta U) = dU + dW = T dS = dQ_{rev}.$$

The change we have made amounts to adding heat dQ_{rev} to the gas which increases its energy by dU and does work dW . Using

$$\frac{1}{\beta} d(N \log Z_o + \beta U) = T dS \tag{3.15}$$

we have

$$\frac{1}{\beta} \propto T$$

and inserting a constant of proportionality k

$$\beta = (kT)^{-1} \tag{3.16}$$

and

$$S = k [N \log Z_o + \beta U] \tag{3.17a}$$

The constant k here is Boltzmann's constant introduced by Planck in 1904 in memory of Boltzmann. Boltzmann established the proportionality of S and $N \log Z_o + \beta U$ but never introduced a constant. In fact, in doing so here we have really made an assumption for (3.15) is only a differential relation. We have really set the additional constant that should appear in (3.17a) equal to zero.

Demonstrating rigorously that (3.17a) holds with the constant equal to zero is difficult and is referred to as Nernst's (1904 -) theorem or the third law of thermodynamics. The demonstration involves showing that the constant is the same for all bodies and phases independent of the energy state spectrum.

For this reason the value of the constant is unimportant. At $T = 0$ when all the systems are in their single ground state ϵ_1 , all bodies irrespective of the state spectrum have this same constant entropy and it is most convenient to set the constant to zero, in which case (3.17a) holds. Excellent discussions of this point have been given by Schrödinger, ter Haar and Wilks.

Equation (3.17a) then leads us directly to Boltzmann's famous relation between S and W_{max} . From (3.10)

$$\begin{aligned} \log W_{max} &= N \log N + \alpha N + \beta U \\ &= N \log N + (\log Z_o - \log N) N + \beta U \\ &= N \log Z_o + \beta U \\ &= \frac{S}{k} \end{aligned}$$

or

$$S = k \log W_{max} \quad (3.18)$$

which is Boltzmann's relation relating the entropy to the number of ways the microscopic state can be formed.

3.4 Correct Boltzmann Counting

Up to this point we have supposed the gas is isolated so that the entropy is the appropriate thermodynamic function to consider. Total isolation is an idealization and some thermal contact between the gas and its surroundings (which we regard as a heat bath) allowing exchange of heat always occurs. In this case, if T is held constant, the most useful thermodynamic function is the Helmholtz free energy, $F(T, V)$. How is F related to Z_o ? For this we go back to

$$S = k [N \log Z_o + \beta U]$$

and

$$F \equiv U - TS = -kT N \log Z_o = -kT \log Z_o^N$$

This may be regarded as the most fundamental link between thermodynamics (through statistical mechanics) and the microscopic properties of the gas reflected in the energy level spectrum ϵ_s - except that it is not quite correct. The $Z_o = \sum_s e^{-\beta \epsilon_s}$ is certainly the correct single particle partition function. However the partition function for the gas is

$$Z = \frac{Z_o^N}{N!} \quad (3.19)$$

and

$$F = -kT \log Z = -kT \log \left(\frac{Z_o^N}{N!} \right) \quad (3.20)$$

We must divide Z_o by $N!$ because in fact the same macroscopic state of the gas is obtained if we permute the particles among the levels (or equivalent among the cells in phase space). It is only the total distribution that counts. For example, the macroscopic state, (e.g. energy, pressure, density) of a gas of helium atoms is unchanged if we permute the atoms about the states with the restriction that the same number n_1 are always in state 1, n_2 in state 2 and so on as before the permutation. This arises since the particles in the gas are indistinguishable in terms of setting the macroscopic state of the gas. The entropy consistent with (3.20) is

$$S = - \left(\frac{\partial F}{\partial T} \right) = k \left[\log \left(\frac{Z_o^N}{N!} \right) + \beta U \right] \quad (3.17b)$$

whereas (3.17a) does not contain the correct counting. Only the free energy and entropy contain the $N!$. In all other thermodynamic functions the $N!$ drops out in differentiation of F . However, we shall see that the $N!$ is required in (3.17b) to make S a properly extensive thermodynamic quantity (See Section 4.4).

In Boltzmann's original formulation of the statistic's for gases, the partition function (or strictly its equivalent) was not divided by $N!$. This led to a famous paradox pointed out later by Gibbs for gases of identical particles. Recognizing the indistinguishability of particles, familiar in quantum mechanics, and dividing Z_o by $N!$ removes Gibbs' paradox. We will also see in examining the Sackur-Tetrode equation that the gas entropy given by (3.17b) is confirmed by experiment while S given by (3.17a) is not.

We might suppose that we could remove this problem by dividing W by $N!$ in (3.6) at the outset. However, doing this does not affect the result $n_s = e^{-\alpha} e^{-\beta \epsilon_s}$ and the same expressions are obtained. We could, however, have divided Z_o in 3.14 by $N!$ at the outset of our connection with the thermodynamics of the gas and this indeed would have led to the correct expression (3.17b) for S . The problem is clearly that in terms of the macroscopic state of the gas, the atoms are indistinguishable. This problem and that of the small cell size plagued Boltzmann's formulation and are gracefully removed in Gibbs' formulation discussed in Section 5.

Eq. (3.20) may be regarded as the fundamental link between the microscopic properties and thermodynamics. The most consistent approach is to calculate F from (3.20) and obtain all other thermodynamic functions by differentiating F , e.g.

$$p = -(\partial F / \partial V)_T, \quad S = -(\partial F / \partial T)_V, \quad U = \partial(\beta F) / \partial \beta$$

If this procedure is followed, we can be sure that the resulting thermodynamic functions will correctly satisfy exactly the usual thermodynamic relations even though an approximate energy spectrum is used. We could, of course, calculate

U directly as

$$U = \frac{N}{Z_o} \sum_s e^{-\beta \epsilon_s} \epsilon_s \quad (3.21)$$

Then, however, U and F, calculated from (3.20), using an approximate energy spectrum might not be exactly thermodynamically consistent. This problem arises since (3.21) and (3.20) can represent different approximations when an approximate energy system is used. If the exact energy spectrum is used then U obtained by differentiating F in (3.20) and obtained from (3.21) are identical.