The Boltzmann distribution and a lesson in logic

Andrew M. Steane

Exeter College and Department of Atomic and Laser Physics, University of Oxford.

February 24, 2025

We want to find the values of p_i which maximise

$$S = -k_{\rm B} \sum_i p_i \ln p_i \tag{1}$$

subject to the constraints

$$\sum_i p_i \epsilon_i = E, \qquad \sum_i p_i = 1. \tag{2}$$

Introduce

$$f \equiv \left(\sum_i \epsilon_i p_i \right) - E \tag{3}$$

$$g \equiv \left(\sum_i p_i \right) - 1. \tag{4}$$

Also introduce Lagrange multipliers $\lambda, , \alpha$ and form

$$y = S + \lambda f + \alpha g. \tag{5}$$

This has a stationary value when

$$\frac{\partial y}{\partial p_j} = 0 \tag{6}$$

where for each j the variables held constant in the partial derivative are all the other $p_{i \neq j}$. Therefore

$$-k_{\rm B} \ln p_j - k_{\rm B} p_j \frac{1}{p_j} + \lambda \epsilon_j + \alpha = 0 \tag{7}$$

which gives

$$k_{\rm B} \ln p_j = \lambda \epsilon_j + (\alpha - k_{\rm B}) \tag{8}$$

hence

$$p_j = A e^{\lambda \epsilon_j / k_{\rm B}} \tag{9}$$

where $A = \exp(\alpha / k_{\rm B} - 1)$. This is the Boltzmann distribution.

There remain two further steps. First we define $\beta = -\lambda / k_{\rm B}$ and so

Boltzmann distribution

$$p_j = A e^{-\beta \epsilon_j} \tag{10}$$

and we define $Z = \sum_i \exp(-\beta \epsilon_i)$. Then the constraint related to g and α is satisfied when

$$p_i = \frac{e^{-\beta \epsilon_i}}{Z}. \tag{11}$$

Finally, we want to know the value of β. One can show that if two systems can exchange energy without a change in their sets of energy levels (hence they are exchanging heat not work) then the entropy of the pair is maximised when they have the same β; this suggests that β is related to temperature. It is not hard to convince oneself that it must be an inverse relationship.

The complete analysis of β is achieved by leaving it as β in the equations obtaining formulae relating U to Z and β, and then F. Eventually we find out that

$$\frac{\partial S}{\partial U} = k_{\rm B} \beta \tag{12}$$

where in the partial derivative all the energy levels ϵ_i are held constant. But thermodynamic temperature T is equal to $\partial U / \partial S$. Hence we deduce that

$$\beta = \frac{1}{k_{\rm B} T}. \tag{13}$$

Comments

It is remarkable how the powerful result (10) emerges so quickly from a few simple statements. I think this method of derivation is conceptually one of the most straightforward. In any derivation one has to start by carefully getting one's head around precisely what the symbols mean and what argument is being employed; in the above I think that process is easier than in other approaches.

1 A logical fallacy

I once had to mark the examination papers for an examination in thermal physics at the University of Oxford. One of the questions asked the student to derive the Boltzmann distribution. Almost the entire cohort gave a faulty derivation, and it was the same fault for all of them. This alerted me to the fact that this was not just a failure in their learning but also a failure in our teaching.

I will now present this faulty argument, in order first to refute it, and then to correct it. The argument can be found in a number of textbooks and other resources. Sometimes it is presented correctly, sometimes not. Beware!

The argument considers a reservoir with a large number of microstates $\Omega(E)$ which is a function of the energy E of the reservoir. We suppose this reservoir exchanges energy with our small system whose energy is ϵ. So in order to conserve energy, when the system has energy ϵ the reservoir has energy $E - \epsilon$. The probability of this state of affairs is therefore

$$p(\epsilon) \propto \Omega(E - \epsilon) \times 1 \tag{14}$$

where the 1 signifies the one state consistent with the energy ϵ of the system.

Now, for $\epsilon \ll E$, by using the Taylor expansion we can always write

$$\begin{aligned} \ln \Omega(E - \epsilon) &= \ln \Omega(E) - \epsilon \frac{\mathrm{d} \ln \Omega}{\mathrm{d} E} + \ldots & (15) \\ &= \ln \Omega(E) - \beta \epsilon + \ldots & (16) \end{aligned}$$

where

$$\beta \equiv \frac{\mathrm{d} \ln \Omega}{\mathrm{d} E} \tag{17}$$

(evaluated at reservoir energy E). This β is a property of the reservoir. In the limit where the further terms in the Taylor expansion are negligible, (16) gives

$$\Omega(E - \epsilon) = \Omega(E) e^{-\beta \epsilon} \tag{18}$$

and therefore, by employing this in (14), we have

$$p(\epsilon) \propto e^{-\beta \epsilon}. \tag{19}$$

So we have the Boltzmann distribution. Or do we? As I warned you, this argument is spurious as it stands. To be fair, it is not completely spurious, but it is incomplete, and without further statements to justify the approximation of dropping the higher order terms it is almost completely useless. Here is why.

What really happened in the above argument is that we took a function of Ω, linearized that function by forming a Taylor series expansion to low order, and then claimed the result to be accurate. But if this were valid then we could equally well pick some other function, say $\tan \Omega$, and linearize that, and then we shall find

$$p(\epsilon) \propto \arctan(A - \alpha \epsilon) \tag{20}$$

for some constants A and α which are properties of the reservoir. (The reader is encouraged to try it.) By this kind of 'reasoning' one can 'derive' that $p(\epsilon)$ is pretty much any function you like.

What is really going is that one finds that $p(\epsilon)$ can be written *to first order approximation* in terms of almost any analytic function, but this says almost nothing about what function $p(\epsilon)$ is really. That information is hiding in the behaviour of all the higher-order terms that were neglected when we took a first order approximation.

I want to underline the importance of this point. Consideration of the higher-order terms here is not just a tidying-up exercise or a kind of afterthought which is not really central to the argument. No: it is almost the entire argument, because until we have done it we can claim only that our formula for $p(\epsilon)$ is valid to first order in ϵ, and nothing more. When we see (19) it is very tempting to think that our work is done and we have derived the Boltzmann distribution. That would be to commit an elementary fallacy of logic. It would be like the following 'argument':

- Major premise. For any integer x, if x is a multiple of 4 then x is a multiple of 3.
- Minor premise. 12 is a multiple of 4.
- Conclusion: 12 is a multiple of 3.

The concluding statement here is a true statement, but the argument is entirely faulty because the major premise is wrong. *The mere fact that the conclusion of an argument is correct does not in itself validate the argument.*

The derivation presented above is not a derivation of the Boltzmann distribution. It is a derivation of the statement

$$p \propto e^{-\beta \epsilon + O(\epsilon^2)}. \tag{21}$$

With all this in mind, we will now present a correct derivation of the Boltzmann distribution. We begin as before, with (14). Then we obtain (16) as before, except that we write the next term explicitly:

$$\ln \Omega(E-\epsilon) \quad = \quad \ln \Omega(E) - \beta\epsilon + \frac{\epsilon^2}{2}\frac{\mathrm{d}\beta}{\mathrm{d}E} + \ldots \tag{22}$$

Note also that by using $\Omega = S/k_{\mathrm{B}}$, where S is the entropy of the reservoir, one finds $\beta = 1/k_{\mathrm{B}}T$. Now

$$\frac{\mathrm{d}\beta}{\mathrm{d}E} = \frac{-1}{k_{\mathrm{B}}T^2}\frac{\partial T}{\partial E} = \frac{-1}{k_{\mathrm{B}}T^2 C} \tag{23}$$

where C is the heat capacity of the reservoir. The condition that this second-order term shall be negligible compared to the first-order term is

$$\frac{\epsilon^2}{2k_{\mathrm{B}}T^2 C} \ll \beta\epsilon \tag{24}$$

which is the requirement

$$C \gg \frac{2\epsilon}{T}. \tag{25}$$

Since we are considering just a single system state at energy ϵ, the quantity on the right here is of order k_{B}, whereas the quantity on the left is of order $N_{\mathrm{R}}k_{\mathrm{B}}$ where N_{R} is the number of particles in the *reservoir*. By taking the size of the reservoir to infinity, we can arrange that the second order term shall vanish completely.

One then needs to look at terms of higher order still. One finds that a sufficient condition for all of them to be negligible is that the reservoir should be large, and with a heat capacity that is not a strong function of energy. In the limit of an infinite reservoir one may then conclude that (19) is exact.

Exercise. To get a really thorough grasp of the above reasoning, it is instructive to see what happens when we don't take the logarithm of Ω, but investigate a Taylor series expansion of Ω itself. Then we have

$$\Omega(E-\epsilon) = \Omega(E) - \epsilon\frac{\mathrm{d}\Omega}{\mathrm{d}E} + \frac{\epsilon^2}{2}\frac{\mathrm{d}^2\Omega}{\mathrm{d}E^2} + \ldots \tag{26}$$

(i) Using thermodynamic reasoning, show that $\mathrm{d}\Omega/\mathrm{d}E = \beta\Omega$.
(ii) Show that the second-order term is

$$\frac{\epsilon^2}{2}\left(-\frac{\Omega}{k_{\mathrm{B}}T^2 C} + \beta^2\Omega\right)$$

and hence that, in the limit $C \to \infty$,

$$p \propto 1 - \beta\epsilon + \frac{1}{2}(\beta\epsilon)^2 + O(\epsilon^3) \tag{27}$$

(iii) Extend the argument to all orders and thus derive (10).
(This exercise is not an efficient way to derive the Boltzmann distribution; it is an exercise in how one must treat a Taylor series if one is aiming to derive an exact result.)