Appendix to Part IV (Optional)

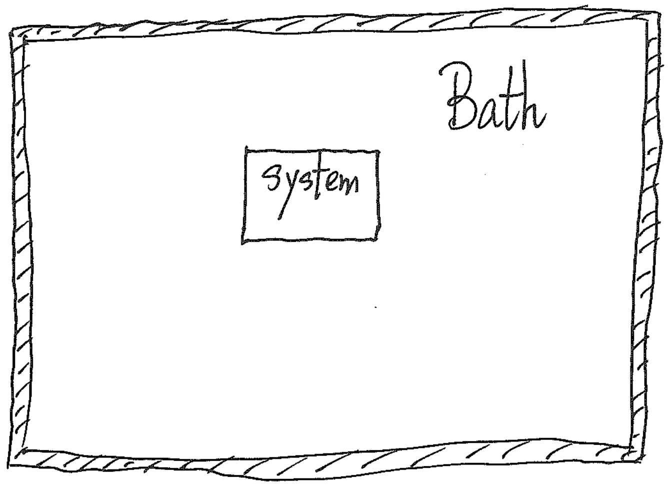
- This appendix gives an alternative derivation to the key result

  P_i = Prob. of finding a system (generally N-particle system) in a state of energy E_i at equilibrium at a temperature T

$$\propto e^{-\frac{E_i}{kT}} = \frac{1}{Z} e^{-\beta E_i} \quad \text{—(1)}$$

  From which, the canonical ensemble theory follows.

- Recall, (1) was derived by considering the set up



- Here, (1) will be derived, again, by considering a big collection of (N-particle in general) systems and using the Lagrange Multipliers method.

  This is why it is given here as an appendix.

- This approach is important in that...
  - it is closer to the spirit of an ensemble theory as Gibbs treated the problem in ~1900
  - it shows that Gibbs' formula of the entropy

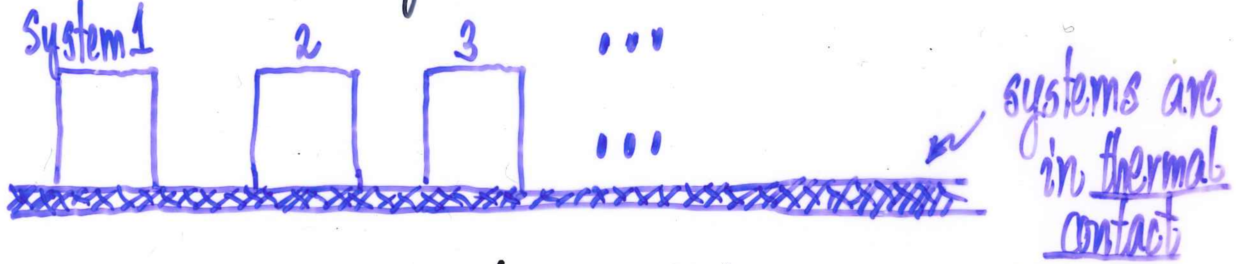
    $$S = -k \sum_i P_i \ln P_i$$

    works for the canonical ensemble.

Appendix A: The Boltzmann distribution (Derivation #2)

Aims:
- an alternative way to get $P_i \propto e^{-\beta E_i}$
- Method of Lagrange multiplers
- connection to the most probable distribution of particles in single-particle states (ch. VII)

- Consider a huge number $\mathcal{N}$ of identical N-particle systems, they are in mutual thermal contact. The whole collection form an isolated system.†



n_r = number of systems in state r (for which the energy is E_r)

E_{tot} = Total energy of the whole ensemble (fixed)

$\therefore$
$$\boxed{\begin{aligned} E_{tot} &= \sum_r n_r E_r = \text{constant} \\ \mathcal{N} &= \sum_r n_r = \text{constant} \end{aligned}}$$
These are constraints.

† We will see that when the collection becomes equilibrium, the collection forms a canonical ensemble.

The scenario is the following: (Microcanonical Ensemble)

- Whole ensemble: Isolated system with E_{tot} fixed
- Wait and wait... members can exchange energy and the whole isolated system approaches equilibrium.
- The whole isolated system's (ensemble) entropy S is a maximum.
- All "microstates" are equally probable.
- What is a microstate?

$\hookrightarrow \left\{ \begin{array}{l} \text{specify system 1's state} \\ \text{specify system 2's state} \\ \vdots \\ \text{specify system } \mathcal{N}\text{'s state} \end{array} \right.$

- Considering all accessible microstates leads to the idea of distributions.
- A distribution: $\{n_1, n_2, \ldots n_r, \ldots\}$

  n_r: # systems in state r

- Since all microstates are equally probable, there is a <u>most probable distribution</u> for which $W(\{n_r\}_{mp})$ is the largest.

  For very large $\mathcal{N}$, $W(\{n_r\}_{mp})$ dominates.

  Thus, if we can find $\{n_r\}_{mp}$ that maximizes W and hence $\ln W$, it is almost always the case that $\{n_1, n_2, \cdots, n_r, \cdots\}_{mp}$ systems in states $1, 2, \cdots, r, \cdots$ of (N,V) system with energy $E_1, E_2, \cdots, E_r, \cdots$

$$\therefore \quad \frac{n_r^{(\text{most probable})}}{\mathcal{N}} = \text{Probability of finding a system to be in a state of energy } E_r$$

$\therefore$ The aim is to find the <u>Most Probable Distribution</u>.

- Find the <u>most probable distribution</u>.

  A distribution is a set of numbers $\{n_r\}$ for distributing $\mathcal{N}$ objects with n_1 in one group, n_2 in another group, and so on.

$$\therefore W(\{n_r\}) = \frac{\mathcal{N}!}{n_1!\, n_2! \cdots} \quad \text{for a distribution.}$$

- The most probable distribution+ maximizes W and hence $\ln W$.

  Assuming all the numbers n_r are $n_r \gg 1$ and using the Stirling approximation:

$$\boxed{\ln W(\{n_r\}) = \mathcal{N} \ln \mathcal{N} - \sum_r n_r \ln n_r}$$

+ Recall:

$$S_{\text{whole ensemble}} = \sum_{\text{all distributions}} k \ln W(\text{distribution})$$

$$\approx k \ln W\left(\{n_r\}^{\text{most probable}}\right)$$

The mathematical problem+ becomes:

Find the set of values of n_r, i.e. $\{n_r\}$, so that

$$\ln W(\{n_r\}) = N \ln N - \sum_r n_r \ln n_r$$

is maximized, under the constraints

$$\sum_r n_r = N = \text{constant}$$

and

$$\sum_r E_r n_r = E_{tot} = \text{constant}$$

- The most convenient way is to use Lagrange's method of undetermined multipliers.
- For the set $\{n_r\}$ we are looking for, $\ln W$ is maximized.

$\therefore$ $\delta(\ln W) = 0$ for all variations of n_r.

$$\delta(\ln W(\{n_r\})) = -\left(\sum_r \delta n_r \ln n_r + \sum_r n_r \frac{1}{n_r}\delta n_r\right) = 0$$

$$\Rightarrow \boxed{-\sum_r (\ln n_r + 1)\,\delta n_r = 0} \qquad \text{(A1)}$$

+ The same problem appears in determining the most probable distribution of particles into single-particle states, as introduced in Ch. III.

- The constraints imply δn_r are NOT independent.

They are related by: $\sum_r n_r = N$ = constant

$$\Rightarrow \boxed{\sum_r \delta n_r = 0} \qquad \text{(A2)}$$

Introduce a yet-to-be-determined Lagrange multiplier+ α,

$$\alpha \sum_r \delta n_r = 0$$

Similarly, $\sum_r E_r n_r = E_{tot}$ = constant

$$\Rightarrow \boxed{\sum_r E_r\,\delta n_r = 0} \qquad \text{(A3)}$$

Introduce a yet-to-be-determined Lagrange multiplier+ β,

$$\beta \sum_r E_r\,\delta n_r = 0$$

Putting (A1), (A2), (A3) together, we have

$$\boxed{-\sum_r (\ln n_r + 1 + \alpha + \beta E_r)\,\delta n_r = 0} \qquad \text{(A4)}$$

+ In general, each constraint carries a multiplier. At this point, β is a parameter. At the end of the derivation and after making contact with physical problems, β can be identified to be $1/kT$.

Math Aside:

not true!

- Look at Eq. (A4), $\boxed{\text{if}}$ δn_r are independent, then we can conclude that $\ln n_r + 1 + \alpha + \beta E_r = 0$ and solve for n_r. Then we are done!
- But the constraints imply δn_r are not independent! So, we cannot jump to conclude $\ln n_r + 1 + \alpha + \beta E_r = 0$.
- Here is the standard argument for the method of Lagrange multipliers. Once you understand it, it is not necessary to repeat the argument every time. Just jump to the conclusion.

We have two constraints:

$$\sum_r \delta n_r = 0 \Rightarrow \delta n_1 + \delta n_2 = -\sum_{r>2} \delta n_r$$

$$\sum_r E_r \delta n_r = 0 \Rightarrow E_1 \delta n_1 + E_2 \delta n_2 = -\sum_{r>2} E_r \delta n_r$$

$\therefore$ We can only take, say, $\{\delta n_3, \delta n_4, \dots\}$ as independent, and then δn_1 and δn_2 are given in terms of $\{\delta n_3, \delta n_4, \dots\}$.

At this point, the multipliers α and β come in: we choose α and β so that

$$\ln n_1 + 1 + \alpha + \beta E_1 = 0$$

and

$$\ln n_2 + 1 + \alpha + \beta E_2 = 0.$$

From Eq. (A4), we have

$$\underbrace{(\ln n_1 + 1 + \alpha + \beta E_1)}_{\overset{\shortparallel}{0}}\delta n_1 + \underbrace{(\ln n_2 + 1 + \alpha + \beta E_2)}_{\overset{\shortparallel}{0}}\delta n_2 + \sum_{r\geq 3} (\ln n_r + 1 + \alpha + \beta E_r)\underbrace{\delta n_r}_{} = 0$$

(r=1 term) forced to vanish by choosing α and β (r=2 term)

δn_r (r=3,4,...) are independent

$$\therefore \ln n_r + 1 + \alpha + \beta E_r = 0 \qquad r \geq 3$$

Thus, from Eq. (A4)

$$\sum_r (\ln n_r + 1 + \alpha + \beta E_r)\delta n_r = 0, \quad \text{(A4)}$$

for this to be true for all δn_r, we must have

$$\boxed{\ln n_r + 1 + \alpha + \beta E_r = 0 \quad \text{for ALL } r}$$

This completes the reasoning behind the method of Lagrange multipliers.

- Go back to Eq. (A4): $\sum_r (\ln n_r + 1 + \alpha + \beta E_r)\delta n_r = 0$,
  it follows that

$$\ln n_r + 1 + \alpha + \beta E_r = 0 \quad , \quad r = 1, 2, \ldots$$

$$\Rightarrow n_r = e^{-1-\alpha} e^{-\beta E_r} \propto e^{-\beta E_r}$$

(Boltzmann distribution)

- The multiplier α is fixed by:

$$n = \sum_r n_r = e^{-1-\alpha} \sum_r e^{-\beta E_r}$$

$$\Rightarrow e^{-1-\alpha} = \frac{n}{\sum e^{-\beta E_r}}$$

$$\therefore n_r = \frac{n\, e^{-\beta E_r}}{\sum_r e^{-\beta E_r}}$$

$$\Rightarrow \boxed{\frac{n_r}{n} = P_r = \frac{e^{-\beta E_r}}{\sum_r e^{-\beta E_r}}}$$

Recall: Physical Meaning of $\frac{n_r}{n}$ is the prob. of finding a system to be in a state of energy E_r

which is just the Boltzmann or canonical distribution, previously derived in Ch. V.

- Making contact with our previous result, it is obvious that $\beta = \frac{1}{kT}$. The denominator $\sum_r e^{-\beta E_r} = Z$ = Partition Function

- Mathematically, the multiplier β is fixed by

$$E_{tot} = \sum_r E_r n_r = \frac{n \sum_r E_r e^{-\beta E_r}}{\sum_r e^{-\beta E_r}}$$

OR

$$\frac{E_{tot}}{n} = \underset{\langle E \rangle}{\text{mean energy per system}} = \frac{\sum_r E_r e^{-\beta E_r}}{\sum_r e^{-\beta E_r}}$$

fixed in ensemble

a familiar formula

formally it is an equation to solve for β

- Since we have also worked on the other derivation, we don't need to go through the process of identification of β as $1/kT$.

This completes derivation #2.

And we know how to use the method of Lagrange multipliers.

Remarks:

(1) Thus, in equilibrium, there are

$\mathcal{N} P_r$ copies of system in state r in the ensemble.

$\therefore$ The ensemble is a canonical ensemble.

(2) In our derivations of P_r, we only assume that the interaction between systems (copies) is weak.

We do NOT need to make any assumption on whether the interaction between particles in the N-particle system is weak or not! Thus,

$$P_r = \frac{1}{Z} e^{-\beta E_r}$$

is a general result, as stressed before.

$\boxed{(3)}$ One may (as in almost all stat. mech. books) apply the result to a system of one particle.

In doing so, it is implicitly assumed that

- the interaction between particles in a N-particle system is WEAK,
- thus, one may consider single-particle states r of energy ε_r

- and $\frac{1}{Z} e^{-\beta \varepsilon_r}$ is the prob. that the single-particle state r is occupied at a temp. T

(We will work on single-particle states later.)

Ref:

See Pathria, "Statistical Mechanics"

Gibbs' Entropy Formula

- We can, of course, do all the physics by the "most probable distribution" method.

Based on the set up, the first quantity to get is

$$S_{\substack{\text{whole collection}\\ \text{of } \mathcal{N} \text{ systems}}} = k \ln W^{(mp)}$$

$$= k \ln \frac{\mathcal{N}!}{n_1! \cdots n_r! \cdots}$$

where $\underbrace{\frac{n_r}{\mathcal{N}} = P_r}_{n_r = \mathcal{N}\cdot P_r} = \frac{1}{Z} e^{-\beta E_r}$

$$\therefore \frac{S_{\text{whole collection}}}{k} = \ln\left(\frac{\mathcal{N}!}{\prod_r n_r!}\right)$$

$$= \mathcal{N}\ln\mathcal{N} - \mathcal{N} - \sum_r (n_r \ln n_r - n_r)$$

$$= \mathcal{N}\ln\mathcal{N} \cancel{-\mathcal{N}} - \sum_r (\mathcal{N}P_r \ln(\mathcal{N}P_r)) + \cancel{\sum_r n_r}$$

$$= \cancel{\mathcal{N}\ln\mathcal{N}} - \cancel{\mathcal{N}\underbrace{\sum_r P_r}_{1\text{ (normalized)}} \ln\mathcal{N}} - \mathcal{N}\sum_r P_r \ln P_r$$

$$= -\mathcal{N}\sum_r P_r \ln P_r$$

$$\therefore \frac{S_{\text{whole collection}}}{\mathcal{N}} = S = \text{Entropy of a system in the collection}$$

$$= -k\sum_r P_r \ln P_r$$

$$\therefore \boxed{S = -k\sum_r P_r \ln P_r} \quad \text{Gibbs Entropy Formula}$$

over all N-particle states

works for the canonical ensemble.

Remarks

- We saw that the formula works for microcanonical ensemble $S = k\ln W = -k\sum_{i=1}^{W}\left(\frac{1}{W}\right)\ln\left(\frac{1}{W}\right)$
- Using the same method, the formula can be shown to work also for the grand canonical ensemble.
- The formula $\frac{S}{k} = -\sum_r P_r \ln P_r = \sum_r P_r(-\ln P_r) \equiv \langle -\ln P_r \rangle$ is the most important formula in information theory.