A. Schekochihin
Stat Mech - HT13
Oxford.

§8. Principle of Maximum Entropy

8.1 Quantifying ignorance

In order to make progress, we will adopt the following, perhaps surprising (in view of your past experience of school and UG physics) attitude to the probabilities $\{p_\alpha\}$. Let us think of them as quantifying our ignorance about the true microstate the system is really in. This of course depends on the information we do possess (or expect to be able to obtain).

- Suppose first that we know nothing at all about the system. Then the only fair way of assigning probabilities to microstates is to suppose that they are all equally likely:

$$\boxed{p_\alpha = \frac{1}{\Omega}} \qquad (1)$$

(in acknowledging our ignorance)

This principle of fairness can be given the status of a postulate, known as the fundamental postulate of SM (aka equal a priori probability postulate [Boltzmann]) ~~[illegible]~~ To wit,

> For an isolated system in equilibrium, all microstates are equally likely

means $\{p_\alpha\}$ are not changing, it's all come to some statistically steady state

this means it is not in contact with anything - which is consistent with us knowing nothing about it

- But this is in fact not an interesting situation for us because the reason we are doing all this is so that we can predict results of measurements. This means we do expect to know something about our system — namely the quantities measured. Those will typically be macroscopic quantities, e.g., mean energy:

$$U = \sum_\alpha p_\alpha E_\alpha \qquad (2)$$

Clearly, any particular measured value of U will be consistent with lots of microstates, so knowing U, while not necessarily consistent with (1), will not constrain the values of p_α's very strongly: indeed there are $\Omega \ggg 1$ p_α's and only one equation (2) that they are required to satisfy. We may have more information in the form of eqns like (2) - but the principle is clear: the amount of info we are ever likely to have (or want) falls hugely short of uniquely fixing every p_α.

- So we would like to find a way of assigning values to $\{p_\alpha\}$ taking into account strictly the info we have and nothing more.

We will adopt the following algorithm (Jaynes §11.4). We have Ω microstates and need to assign $p_1 \dots p_\Omega$ probabilities to them, subject to $\sum_\alpha p_\alpha = 1$.

Choose some $\mathcal{N} \gg \Omega$ and embark on assigning $\mathcal{N}$ "quanta" of probability (each of magnitude $\frac{1}{\mathcal{N}}$) to Ω microstates [imagine tossing $\mathcal{N}$ pennies into Ω boxes in an equiprobable way]

After we have used up all $\mathcal{N}$ quanta, suppose we find $\mathcal{N}_1$ quanta in microstate 1

$\mathcal{N}_2$ " " " 2

...

$\mathcal{N}_\alpha$ " " " α

...

$\mathcal{N}_\Omega$ " " " Ω

which corresponds to assignment of probabilities

$$p_\alpha = \frac{\mathcal{N}_\alpha}{\mathcal{N}}, \quad \alpha = 1, 2 \ldots \Omega \tag{3}$$

Now check whether this set $\{p_\alpha\}$ satisfies the constraint(s) imposed by the available information, e.g. (2). If it does not, reject this assignment of probabilities and repeat the experiment. Keep going until a satisfactory set $\{p_\alpha\}$ is found.

What is the most likely outcome of this game? The # of ways in which an assignment (3) can obtain is the # of ways of choosing $\mathcal{N}_1 \ldots \mathcal{N}_\Omega$ quanta out of a set of $\mathcal{N}$:

$$W = \frac{\mathcal{N}!}{\mathcal{N}_1! \ldots \mathcal{N}_\Omega!} \tag{4}$$

All outcomes are equiprobable, so the most likely assignment $\{\mathcal{N}_\alpha\}$ is the one that maximizes W subject to constraints imposed by the available info.

[It is possible to prove that this maximum is very sharp for large $\mathcal{N}$.]

Note that we were at liberty to choose $\mathcal{N}$ as large as we liked and so we may assume that all $\mathcal{N}_\alpha \gg 1$ and use <u>Stirling's formula</u> to evaluate factorials:

$$\ln \mathcal{N}_\alpha ! = \mathcal{N}_\alpha \ln \mathcal{N}_\alpha - \mathcal{N}_\alpha + O(\sqrt{\mathcal{N}_\alpha}) \qquad (5)$$

So

$$\ln W = \underbrace{\mathcal{N}}_{\sum_\alpha \mathcal{N}_\alpha} \ln \mathcal{N} - \cancel{\underbrace{\mathcal{N}}_{\sum_\alpha \mathcal{N}_\alpha}} + O(\sqrt{\mathcal{N}}) - \sum_\alpha \left[\mathcal{N}_\alpha \ln \mathcal{N}_\alpha - \cancel{\mathcal{N}_\alpha} + O(\sqrt{\mathcal{N}_\alpha}) \right]$$

$$= - \sum_\alpha \mathcal{N}_\alpha \ln \frac{\mathcal{N}_\alpha}{\mathcal{N}} + O(\sqrt{\mathcal{N}}) =$$

$$= - \mathcal{N} \sum_\alpha p_\alpha \ln p_\alpha + O(\sqrt{\mathcal{N}})$$

or, more precisely, if $\mathcal{N}, \mathcal{N}_\alpha \to \infty$ as $\frac{\mathcal{N}_\alpha}{\mathcal{N}} \to p_\alpha = \text{const}$,

$$\frac{1}{\mathcal{N}} \ln W \to \boxed{- \sum_\alpha p_\alpha \ln p_\alpha \equiv S_G} \quad \text{Gibbs entropy}, \qquad (6)$$

or <u>information entropy</u> (Shannon) ("amount of ignorance")

Maximising W is the same as maximising S_G, so the import of this quantity is that <u>the "fairest" assignment of probabilities $\{p_\alpha\}$ subject to some info will correspond to the maximum of S_G, subject to constraints imposed by that info.</u>

<u>NB</u>: 1) Since $0 < p_\alpha \le 1$, $S_G \ge 0$

($p_\alpha = 0$ means α is not an allowed state

$p_\alpha = 1$ means there's only one state and $S_G = 0$)

we have perfect knowledge $\Leftrightarrow$ 0 ignorance

2) S_G depends only on the probabilities $\{p_\alpha\}$, <u>not</u> on the quantum numbers (random variables) these probabilities describe (e.g. E_α). This means that no change of variables (e.g. $E_\alpha \to f(E_\alpha)$) can affect S_G.

S_G measures the amount of ignorance associated with the probability distribution $\{p_\alpha\}$:

3) The # of all possible distributions of $\mathcal{N}$ probability quanta over Ω microstates is $\Omega^{\mathcal{N}}$, which is clearly the max. value W could take:

$$W_{max} = \Omega^{\mathcal{N}}$$

$$\frac{1}{\mathcal{N}} \ln W_{max} = \ln \Omega = S_{G,max}$$ the max value S_G can ever have

If $p_\alpha = \frac{1}{\Omega}$ (everything is equally likely, <u>ignorance is total</u>),

$$S_G = -\sum_\alpha \frac{1}{\Omega} \ln \frac{1}{\Omega} = \ln \Omega = S_{G,max} \quad (7)$$

[and if $p_\alpha = 1$, $S_G = 0$, <u>complete knowledge</u>]

<u>Information content</u> of a distribution is

$$I(p_1, \ldots, p_\Omega) = S_{G,max} - S_G(p_1, \ldots, p_\Omega)$$
$$= \ln \Omega + \sum_\alpha p_\alpha \ln p_\alpha \quad (8)$$

(Shannon 1948)

Maximising S_G is the same as minimising I.

<u>Read</u> C. Shannon, Bell Systems Tech. Journal <u>27</u>, 379 (1948)

(I will discuss this a little more later.)

L.1 ended here.

8.2 Lagrange Multipliers

A mathematical digression: how do we maximize a function, say $S_G(p_1, \dots, p_\Omega)$, subject

Ω variables

to some constraint of general form

$$F(p_1, \dots, p_\Omega) = 0 \qquad (9)$$

$\hookrightarrow$ e.g., $\sum_\alpha p_\alpha E_\alpha - U = 0$ eq. (2) ?

At the point of maximum (or, more precisely, extremum),

$$dS_G = \frac{\partial S_G}{\partial p_1} dp_1 + \dots + \frac{\partial S_G}{\partial p_\Omega} dp_\Omega = 0, \qquad (10)$$

but $\{dp_\alpha\}$ are not independent because $\{p_\alpha\}$ are only allowed to be changed subject to (9), i.e., they have to stay on the surface $F(p_1 \dots p_\Omega) = 0$.

So F cannot change:

$$dF = \frac{\partial F}{\partial p_1} dp_1 + \dots + \frac{\partial F}{\partial p_\Omega} dp_\Omega = 0 \qquad (11)$$

This determines one of the dp_α's in terms of others: e.g.,

$$dp_\Omega = -\left(\frac{\partial F/\partial p_1}{\partial F/\partial p_\Omega} dp_1 + \dots + \frac{\partial F/\partial p_{\Omega-1}}{\partial F/\partial p_\Omega} dp_{\Omega-1} \right) \qquad (12)$$

Sub. (12) into (10):

$$dS_G = \left(\frac{\partial S_G}{\partial p_1} - \frac{\partial S_G/\partial p_\Omega}{\partial F/\partial p_\Omega} \frac{\partial F}{\partial p_1} \right) dp_1 + \dots$$

$$\dots + \left(\frac{\partial S_G}{\partial p_{\Omega-1}} - \underbrace{\frac{\partial S_G/\partial p_\Omega}{\partial F/\partial p_\Omega}}_{\equiv\ \lambda} \frac{\partial F}{\partial p_{\Omega-1}} \right) dp_{\Omega-1} = 0 \qquad (13)$$

In (13), $dp_1 \ldots dp_{\Omega-1}$ are now all independent, so (13) implies

$$\left.\begin{aligned} &\frac{\partial S_G}{\partial p_\alpha} - \lambda \frac{\partial F}{\partial p_\alpha} = 0 \quad \text{for } \alpha = 1, \ldots, \Omega-1 \\ &\text{where, by definition of } \lambda, \\ &\frac{\partial S_G}{\partial p_\Omega} - \lambda \frac{\partial F}{\partial p_\Omega} = 0 \end{aligned}\right\} \quad (14)$$

So, we now have $\Omega+1$ unknowns, $p_1 \cdots p_\Omega, \lambda$, and $\Omega+1$ equations for them: (14) and (9)!

$$\begin{cases} \dfrac{\partial S_G}{\partial p_\alpha} - \lambda \dfrac{\partial F}{\partial p_\alpha} = 0 \quad \text{for } \alpha = 1, \ldots, \Omega & (14) \\ F(p_1 \cdots p_\Omega) = 0 & (9) \end{cases}$$

But this is exactly the equations that we would get if we wanted to <u>maximize</u> $S_G - \underbrace{\lambda}_{\text{Lagrange multiplier}} F$ wrt $p_1 \cdots p_\Omega, \lambda$ <u>and with no constraints</u>:

$$d(S_G - \lambda F) = \sum_\alpha \left(\frac{\partial S_G}{\partial p_\alpha} - \lambda \frac{\partial F}{\partial p_\alpha} \right) dp_\alpha - F\, d\lambda = 0$$

implies

NB: If there are several constraints

$$F_i(p_1 \cdots p_\Omega) = 0 \qquad i = 1, \ldots, m \qquad (15)$$

the above procedure generalizes trivially to maximizing

$$S_G - \sum_i \lambda_i F_i \quad \text{wrt } p_1, \ldots, p_\Omega, \lambda_1, \ldots, \lambda_m \qquad (16)$$

(as many Lagrange multipliers as there are constraints)

Obviously, we must have $m < \Omega$ (but we will always have fewer constraints than microstates!)

Let us check that this works in the case where we know the answer: $\boxed{\text{8.3 Isolated system}}$

We know nothing. So the only constraint is

$$\sum_\alpha p_\alpha = 1$$

We need

$$S_G(p_1 \cdots p_\Omega) \to \max \quad \text{subject to } \sum_\alpha p_\alpha = 1 \qquad (17)$$

$$\Updownarrow$$

$$S_G - \lambda\left(\sum_\alpha p_\alpha - 1\right) \to \max \text{ unconditionally}$$

$$dS_G - \lambda \sum_\alpha dp_\alpha - \left(\sum_\alpha p_\alpha - 1\right) d\lambda = 0$$

$$\|$$

$$-\sum_\alpha \left(\ln p_\alpha \, dp_\alpha + p_\alpha \cdot \frac{1}{p_\alpha} dp_\alpha\right)$$

$$-\sum_\alpha dp_\alpha \left(\underbrace{\ln p_\alpha + 1 + \lambda}_{=0}\right) - \underbrace{\left(\sum_\alpha p_\alpha - 1\right)}_{=0} d\lambda = 0$$

$$\Downarrow$$

$$p_\alpha = e^{-(1+\lambda)} \longrightarrow \sum_\alpha e^{-(1+\lambda)} = 1$$

$$\Omega e^{-(1+\lambda)} = 1$$

$$e^{-(1+\lambda)} = \frac{1}{\Omega}$$

$$\boxed{p_\alpha = \frac{1}{\Omega}} \quad \text{everything is equiprobable, as expected.} \qquad (18)$$

$$S_G = \ln \Omega \quad \text{max. ignorance.} \qquad (19)$$

OK, so the method works!