

## Thermodynamics and information

Consider an arbitrary random variable  $x$  which can take on  $d$  different levels  $x_0, x_1, \dots, x_{d-1}$ . We don't know what are the probabilities  $P_m = P(x = x_m)$ . But we know the expectation value of some function  $g(x)$

$$\langle f(x) \rangle = \sum_m P_m f(x_m) = G \quad (1)$$

where  $G$  is some number.

Can we figure out what is  $P_m$  only from this fact? Of course not! We only know one number,  $G$ , so there is definitely not enough information.

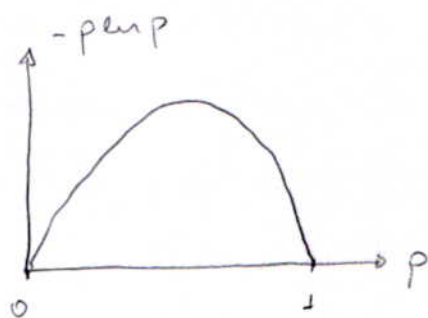
But let's ask this from the point of view of statistical inference: what is the best guess we can make about  $P_m$  given that all we know is that  $\langle g(x) \rangle = G$ . The answer to this was given by in parts by Shannon in 1948 and then Jaynes in 1957.

Shannon showed that there is a unique function for quantifying the amount of uncertainty in a distribution  $P_m$ , which is given by

$$S = - \sum_m P_m \ln P_m \quad (2)$$

this is called the Shannon entropy and is the starting point of information theory.

The entropy quantifies our lack of information about  $x$ .  
 It is useful to remember the behavior of the function  $-p \ln p$



It is zero whenever we have certainty; that is, when  $p=0$  or  $p=1$ .  
 Otherwise, it is always positive.

We thus see that Eq (2) is a sum of positive contributions,  
 which are larger for  $P_n$ 's far from 0 or 1. The maximum of  
 $S$  occurs for a uniform distribution

If  $P_n = \frac{1}{d}$  then  $S = \ln d$

(3)

Conversely, if  $P_e = 1$  for some  $e$ , and all other  $P_n$ 's are zero,  
 then  $S = 0$ . Thus, we have:

Minimum  $S = 0$  : deterministic case (4)  
 Maximum  $S = \ln d$  : uniform distribution  
 (maximum ignorance)

We can also define the information about  $x$

$$I = \ln(d) - S \tag{5}$$

It goes from  $\ln(d)$ , which is the largest amount of information  
 we may have, to 0 in the uniformly distributed case.

Let's now go back to our original problem: we have some rv  $X$ , for which we don't know  $P_m$ . Instead, all we know is that  $\langle g(x) \rangle = G$ . What is the best (i.e., least biased) distribution  $P_m$  that agrees with this information?

The answer is given by the principle of maximum entropy, or MaxEnt: the distribution  $P_m$  is that which maximizes  $S$  subject to the constraint  $\langle g(x) \rangle = G$

To carry out this constrained maximization we introduce Lagrange multipliers, by defining

$$\tilde{S} = - \sum_m P_m \ln P_m + \alpha (1 - \sum_m P_m) + \beta (G - \sum_m g(x_m) P_m) \quad (6)$$

The Lagrange multiplier  $\alpha$  ensures that the probabilities are normalized

$$\frac{\partial \tilde{S}}{\partial \alpha} = 1 - \sum_m P_m = 0 \implies \sum_m P_m = 1 \quad (7)$$

Similarly,  $\beta$  ensures that  $\langle g(x) \rangle = G$

$$\frac{\partial \tilde{S}}{\partial \beta} = G - \sum_m g(x_m) P_m = 0 \implies \sum_m g(x_m) P_m = G \quad (8)$$

then we maximize  $\tilde{S}$  with respect to each  $P_m$ :

$$\frac{\partial \tilde{S}}{\partial P_m} = - \ln P_m - \frac{P_m}{P_m} - \alpha - \beta g(x_m) = 0$$

thus

$$\ln P_m = -1 - \alpha - \beta g(x_m)$$

or

$$P_m = e^{-\alpha-1} e^{-\beta g(x_m)} \quad (9)$$

We can call  $e^{-\alpha-1} = 1/2$ , since this is just a constant, thus we conclude the distribution which maximizes the entropy subject to  $\langle g(x) \rangle = G$  is given by

$$P_m = \frac{e^{-\beta g(x_m)}}{2} \quad (10)$$

-12

Side comment: I forgot to say, but the MaxEnt also applies to the case where we have no extra information. Then there is only one Lagrange multiplier

$$\bar{S} = - \sum_m P_m \ln P_m + \alpha (1 - \sum_m P_m)$$

and the maximization gives

$$\frac{\partial \bar{S}}{\partial P_m} = - \ln P_m - 1 - \alpha = 0$$

thus, all  $P_m$  must be equal. By normalization this then implies

$$P_m = \frac{1}{d}$$

which is the uniform distribution. The entropy in this case achieves its largest value possible

$$S = \ln(d)$$

thus, the conclusion is that if we know nothing about  $X$ , then the best possible guess we can make is a uniform distribution (maximum ignorance)

## Entropy and free energy

The connection with statistical mechanics should now be a bit evident. We identify the Gibbs state of a system as the state which maximizes the entropy subject to the constraint that

$$U = \langle H \rangle = \sum_m E_m P_m = \mathcal{E} \quad (11)$$

for some value of  $\mathcal{E}$

$$\bar{S} = - \sum_m P_m \ln P_m + \alpha (1 - \sum_m P_m) + \beta (\mathcal{E} - \sum_m E_m P_m) \quad (12)$$

which will result in

$$P_m = \frac{e^{-\beta E_m}}{Z} \quad \beta = \frac{1}{T} \quad (13)$$

We thus see that the structure of the Gibbs state may be found solely from statistical inference. If you are interested in learning more about this, check out E. Jaynes, *Physical Review*, 106, 620-630 (1957).

A comment, by the way: the Shannon entropy, Eq (2), also appeared (way before Shannon) in the works of Gibbs. So the fact that the same quantity appears in both statistical mechanics and information theory reinforces the deep connection between the two approaches.

Going back to (12), let us now define a new quantity, which is called free energy (the origin of the name "free" will be clarified later)

$$F = \langle U \rangle - TS \quad (14)$$

or

$$F = \sum_m (E_m P_m + T P_m \ln P_m) \quad (15)$$

Then (12) may be written as

$$\tilde{Z} = -\beta F + \alpha (1 - \sum_m P_m)$$

Thus we see that the following two statements are equivalent:

1) Equilibrium is the state which maximizes the entropy subject to  $\langle H \rangle = U$

2) Equilibrium is the state which minimizes the free energy

this renders a really cool interpretation to the free energy. At zero temperature, equilibrium would be the state of smallest energy (the ground state). But for  $T \neq 0$  we also have some disorder. So equilibrium is not the state which minimizes  $\langle H \rangle$ . Instead, there is a competition between  $\langle H \rangle$  and  $-TS$  and so equilibrium is instead the state which minimizes the free energy

## Thermodynamics

In equilibrium the  $P_m$  are given by the Gibbs distribution (13). Then writing

$$\ln P_m = -\beta E_m - \ln z \quad (17)$$

and substituting in (15), we find

$$\boxed{F = -T \ln z} \quad \boxed{(z = e^{-\beta F})} \quad (18)$$

thus we see that the partition function is more than just a normalization constant, but actually has a physical meaning.

In fact, it turns out, all thermodynamic quantities are actually encoded in  $z$ . For instance, start with

$$z = \sum_m e^{-\beta E_m} \quad (19)$$

and differentiate with respect to  $\beta$

$$\frac{\partial z}{\partial \beta} = - \sum_m E_m e^{-\beta E_m}$$

we thus see that

$$\frac{1}{z} \frac{\partial z}{\partial \beta} = - \frac{\sum_m E_m e^{-\beta E_m}}{\sum_m e^{-\beta E_m}} = -U$$

Hence, from  $z$  we can compute the internal energy as

$$\boxed{U = \langle H \rangle = - \frac{\partial}{\partial \beta} \ln z} \quad (20)$$

Comparing now (18) and (20) we get

immersion

$$U = \frac{\partial}{\partial \beta} \left( \frac{F}{T} \right)$$

using

$$\frac{\partial}{\partial \beta} = \frac{\partial T}{\partial \beta} \frac{\partial}{\partial T} = -T^2 \frac{\partial}{\partial T} \quad (21)$$

we get

$$U = -T^2 \frac{\partial}{\partial T} \left( \frac{F}{T} \right) \quad (22)$$

And finally, once we have  $F$  and  $U$ , then it's easy to compute the entropy since

$$F = U - TS \quad (23)$$

so  $S = (U - F)/T$ . Or, expanding (22),

$$U = F - T \frac{\partial F}{\partial T} \quad (24)$$

thus

$$S = \frac{U - F}{T} = - \frac{\partial F}{\partial T} \quad (25)$$



## Heat capacity (specific heat)

Another important quantity, which is one of the most measured quantities in the lab, is the heat capacity or specific heat (there is actually a difference between these two names, which I will explain in a second). It is defined as the sensitivity of the system's internal energy, to changes in temperature

$$C = \frac{\partial U}{\partial T} \quad (26)$$

Since we set  $k_B = 1$ , for us  $C$  is dimensionless. Otherwise  $C$  would have units of  $k_B$ . Experimental data is usually given in  $J/mol K$ . To convert (26) to these units, multiply it by the gas constant

$$R = 8.314 J/mol K \quad (27)$$

Using Eq (24) we may also write  $C$  as

$$C = \frac{\partial F}{\partial T} - \frac{\partial}{\partial T} \left[ T \frac{\partial F}{\partial T} \right] = \frac{\partial F}{\partial T} - \frac{\partial F}{\partial T} - T \frac{\partial^2 F}{\partial T^2}$$

thus

$$C = \frac{\partial U}{\partial T} = - T \frac{\partial^2 F}{\partial T^2} \quad (28)$$