

Entropy + Information

In canonical ensemble we had

$$\text{prob to be in energy } E \quad P(E) = \frac{\Omega(E) e^{-E/k_B T}}{\Delta Q_N}$$

or if we label microstates by an index i then the prob to be in state i is

$$P_i = \frac{e^{-E_i/k_B T}}{Q_N} \quad \text{where } Q_N = \sum_i e^{-E_i/k_B T}$$

Consider the average value of $\ln P_i$

$$\langle \ln P_i \rangle = \sum_i P_i \ln P_i \quad \text{by definition of average}$$

$$\text{But also } \langle \ln P_i \rangle = \left\langle \ln \left[\frac{e^{-E_i/k_B T}}{Q_N} \right] \right\rangle$$

$$= -\frac{\langle E \rangle}{k_B T} - \ln Q_N$$

$$\Rightarrow k_B T \langle \ln P_i \rangle = -\langle E \rangle + A = -T \langle S \rangle$$

Entropy as computed in canonical ensemble

$$\text{as } A = E - TS$$

$$\Rightarrow \boxed{\langle S \rangle = -k_B \sum_i P_i \ln P_i}$$

where P_i is the prob to be in state i

Note: above was derived for canonical ensemble.

But it also holds for the microcanonical ensemble.

In microcanonical, the prob to be in state i is $1/\Omega(E)$

for a state with $E_i = E$, and zero otherwise. Equally likely to be in any state with energy E

$$\Rightarrow -k_B \sum_i p_i \ln p_i = -k_B \sum_i \left(\frac{1}{\Omega}\right) \ln\left(\frac{1}{\Omega}\right)$$

↑ sum over only states in energy shell about E .

But the terms in the sum are all equal, and the number of terms is just the number of states at energy E , i.e. Ω .

$$\begin{aligned} \Rightarrow -k_B \sum_i \left(\frac{1}{\Omega}\right) \ln\left(\frac{1}{\Omega}\right) &= -k_B \left(\frac{1}{\Omega}\right) \ln\left(\frac{1}{\Omega}\right) \sum_i 1 \\ &= -k_B \left(\frac{1}{\Omega}\right) \ln\left(\frac{1}{\Omega}\right) (\Omega) = -k_B \ln\left(\frac{1}{\Omega}\right) \\ &= k_B \ln \Omega \end{aligned}$$

$$\text{So } -k_B \sum_i p_i \ln p_i = k_B \ln \Omega = S(E) \quad \text{entropy in microcanonical ensemble}$$

$$\text{So } \boxed{S = -k_B \sum_i p_i \ln p_i} \quad \text{works for both microcanonical and canonical ensembles!}$$

~~Shannon~~

Shannon (1948) turned this relation backwards, in developing a close relation between entropy and information theory.

Consider a system with states labeled by i , and P_i is the probability for the system to be in state i .

We want to define a measure of how disordered the distribution P_i is. Call this disorder measure S (it will turn out to be the entropy). The bigger (smaller) S is, the more (less) disordered the system is, the less (more) information we have about the probable state of the system.

We want S to satisfy the following properties

1) If $P_j = \begin{cases} 1 & j=i \\ 0 & j \neq i \end{cases}$ then the state of the system is exactly known to be i . This should have $S=0$ as there is no uncertainty, no disorder

2) For equally likely P_i , i.e. all $p_i = 1/N$ for N states, the system is maximally disordered, i.e. S is max. possible value for all possible N state distributions.

3) S should be additive over independently random systems.

To explain what we mean by (3),
suppose we have one system with N equally likely
states labeled by $n=1, \dots, N$, and a second system with
 M equally likely states labeled by $m=1, \dots, M$.

The combined system ~~has~~ ^{has} $N \times M$ equally states labeled
by the pairs (n, m) . We want

$$S(N \times M) = S(N) + S(M)$$

The function with this property is the logarithm. We
use the natural log, although any base would do.

⇒ For a system of N equally likely states,

$$S = k \ln N \quad \text{where } k \text{ is an arbitrary} \\ \text{proportionality constant.}$$

(Note: if we take $k = k_B$ then above is same as the
definition of entropy in the microcanonical ensemble!)

Suppose that all states are not equally likely.
What is S in such a case?

Consider a system which has two possible states
1 and 2. The prob to be in 1 is p_1 . The
prob to be in 2 is $p_2 = 1 - p_1$. In general $p_1 \neq p_2$,
i.e. the states need not be equally likely.

What is the disorder of this two state system, $S(p_1, p_2)$?

Consider N copies of the two state system.

By additivity of S we want the disorder of this joint system of N copies to be

$$(*) \quad S = N S(p_1, p_2)$$

Now in any given sample of the N copy system, M of the systems will be in state 1, while $N-M$ are in state 2. The prob for this will be given by the binomial distribution

$$P_M = \frac{N!}{M!(N-M)!} p_1^M p_2^{N-M} \leftarrow \text{prob} \left\{ \begin{array}{l} M \text{ in state 1} \\ (N-M) \text{ in state 2} \end{array} \right\}$$

For N large, this probability is very strongly peaked about the average $M = Np_1$. We have

average # systems in state 1 $\langle n_1 \rangle = Np_1$

standard deviation of # in state 1 $\sqrt{\langle n_1^2 \rangle - \langle n_1 \rangle^2} = \sqrt{Np_1p_2}$

so relative width of distribution is $\frac{\sqrt{\langle n_1^2 \rangle - \langle n_1 \rangle^2}}{\langle n_1 \rangle} \sim \frac{1}{\sqrt{N}}$

$\rightarrow 0$ as $N \rightarrow \infty$.

\Rightarrow as N gets large we almost always find the system of N copies with Np_1 in state 1 and Np_2 in state 2.

How many ways are there to choose which ^{Np_1} of the N two level sub-systems are in state 1?

There are $\frac{N!}{(Np_1)!(Np_2)!}$ ways $(Np_2 = N(1-p_1))$

each of these ways are equally likely!

\Rightarrow the entropy of the N copy system is

$$S = k \ln \left[\frac{N!}{(Np_1)!(Np_2)!} \right] \quad \text{log of \# equally likely states!}$$

$$= k \left[\ln N! - \ln(Np_1)! - \ln(Np_2)! \right]$$

use Stirling formula

$$= k \left[N \ln N - N - Np_1 \ln Np_1 + Np_1 - Np_2 \ln Np_2 + Np_2 \right]$$

use $Np_1 + Np_2 = N$ as $p_1 + p_2 = 1$

$$= k N \left[\ln N - p_1 \ln N - p_1 \ln p_1 - p_2 \ln N - p_2 \ln p_2 \right]$$

$$\Rightarrow S = k N \left[-p_1 \ln p_1 - p_2 \ln p_2 \right] \quad \text{since } p_1 + p_2 = 1$$

But by (*) we expect $S = N S(p_1, p_2)$

$$\Rightarrow S(p_1, p_2) = -k \left[p_1 \ln p_1 + p_2 \ln p_2 \right]$$

Similarly, if our system had m possible states, with probabilities p_1, p_2, \dots, p_m , and we took N copies of this m level system, the joint system would have Np_1 subsystems in state 1, Np_2 in state 2, \dots , Np_m in state m . The number of equally likely ways to divide the N subsystems the way is

$$\frac{N!}{(Np_1)!(Np_2)! \dots (Np_m)!}$$

And so a similar line of argument results in

$$S(p_1, \dots, p_m) = -k [p_1 \ln p_1 + p_2 \ln p_2 + \dots + p_m \ln p_m]$$

$$S(\{p_i\}) = -k \sum_i p_i \ln p_i$$

↑
Defines our measure of the disorder of the prob distribution p_i . We see it agrees with what we found for the entropy in both canonical and microcanonical ensembles.

But now we will use it to derive the microcanonical and the canonical ensembles!

S above agrees with the desired properties (1) and (2).

$S=0$ if any $p_i=1$ and all others are zero.

We soon see that S is max if all p_i are equal.

We can now use the above as our definition of entropy and define equilibrium as the prob distribution that maximizes S , subject to whatever constraints may exist on the distribution. Each such constraint represents some "information" we have about the system.

microcanonical ensemble - each state i has an energy E_i

We have $p_i = 0$ for $E_i \neq E$, $p_i \neq 0$ for $E_i = E$

Considering only those states i with $E_i = E$, we now want to maximize S over these non-zero p_i .

We want to maximize $S = -k \sum_i p_i \ln p_i$

subject to the constraint $\sum_i p_i = 1$ (normalization of probabilities)

Use method of Lagrange multipliers

\Rightarrow maximize in an unconstrained way

$$S + k\lambda \sum_i p_i$$

where λ is the Lagrange multiplier - we then determine the value of λ by imposing the constraint. So if there are N states of energy E , the maximization gives

$$0 = \frac{\partial}{\partial p_i} (S + k\lambda \sum_i p_i) = \frac{\partial}{\partial p_i} \left(-k \sum_j (p_j \ln p_j - \lambda p_j) \right)$$

$$\Rightarrow p_i \left(\frac{1}{p_i} \right) + \ln p_i - \lambda = 0$$

$$1 + \ln p_i - \lambda = 0$$

$$p_i = e^{\lambda-1} \quad \text{same for all } i$$

⇒ distribution that maximizes S is equally likely states

$$\sum_i p_i = N e^{\lambda-1} = 1 \Rightarrow \lambda = 1 + \ln(N) = 1 - \ln N$$

$$\Rightarrow p_i = e^{\lambda-1} = e^{-\ln N} = \frac{1}{N} \text{ as expected}$$

⇒ in microcanonical ensemble at energy E , all states with energy E are equally likely.

Canonical Ensemble

Now any E_i is allowed, but we have the constraint that the average energy $\langle E \rangle$ is fixed $\Rightarrow \sum_i p_i E_i = \langle E \rangle$ is fixed. We still have the constraint that

$\sum_i p_i = 1$. Thus the maximization requires two Lagrange multipliers.

$$0 = \frac{\partial}{\partial p_i} \left(-k \sum_j \left[p_j \ln p_j - \lambda p_j + \beta p_j E_j \right] \right)$$

$$\Rightarrow 0 = 1 + \ln p_i - \lambda + \beta E_i$$

$$p_i = e^{\lambda-1} e^{-\beta E_i}$$

$$\text{Normalization} \Rightarrow \sum_i p_i = e^{\lambda-1} \sum_i e^{-\beta E_i} = 1$$

$$\Rightarrow e^{\lambda-1} = \frac{1}{\sum_i e^{-\beta E_i}}$$

$$\Rightarrow p_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}$$

Determine β by condition that

$$\sum_i \frac{e^{-\beta E_i} E_i}{e^{-\beta E_i}} = \langle E \rangle \text{ fixed average energy}$$

If we interpret $\beta = \frac{1}{k_B T}$, we recover the canonical distribution!

More generally if we had any quantity X constrained, i.e. X_i is value in state i , and average value

$$\langle X \rangle = \sum_i p_i X_i \text{ is fixed, then}$$

$$p_i = \frac{e^{-\beta X_i}}{\sum_j e^{-\beta X_j}} \text{ gives maximum } S \text{ consistent with the constraint.}$$

$$\beta \text{ determined by requiring } \frac{\sum_i X_i e^{-\beta X_i}}{\sum_j e^{-\beta X_j}} = \langle X \rangle$$

gives the desired value of $\langle X \rangle$.

We can use the definition

$$S = -k_B \sum_i p_i \ln p_i$$

more generally than for systems in equilibrium in the thermodynamic limit. It can be used just as well for systems of finite size, and for systems out of equilibrium.