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, ie all $p_i = 1/N$ for N states, the system is maximally disordered, ie S is max possible value for all possible N state distributions.

3) S should be <u>additive</u> over independently random systems.

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

The combined system 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.

$\Rightarrow$ For a system of N equally likely states,

$$S = k \ln N$$

where k is an arbitrary 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 $p_1 \neq p_2$, ie 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

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

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

$$P_M = \frac{N!}{M!\,(N-M)!}\, p_1^{M}\, p_2^{N-M} \quad \leftarrow \text{prob} \begin{Bmatrix} M \text{ in state } 1 \\ (N-M) \text{ in state } 2 \end{Bmatrix}$$

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

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

standard deviation of # in state 1 $\quad \sqrt{\langle n_1^2\rangle - \langle n_1\rangle^2} = \sqrt{N p_1 p_2}$

so relative width of distribution is $\dfrac{\sqrt{\langle n_1^2\rangle - \langle n_1\rangle^2}}{\langle n_1\rangle} \sim \dfrac{1}{\sqrt{N}}$
$\to 0$ as $N \to \infty$.

$\Rightarrow$ 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 $\dfrac{N!}{(Np_1)!\,(Np_2)!}$ ways $\quad (Np_2 = N(1-p))$

each of these ways are <u>equally likely</u>!

$\Rightarrow$ the entropy of the N copy system is

$$S = k \ln\left[\frac{N!}{(Np_1)!\,(Np_2)!}\right]$$ 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 $p_1 + p_2 = 1$

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

$$\Rightarrow S = kN\left[-p_1 \ln p_1 - p_2 \ln p_2\right]$$ since $p_1 + p_2 = 1$

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

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

Similarly, if our system had m possible states, with probabilities $p_1, p_2, \ldots, p_m$, 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, $\ldots$, Np_m in state m. The number of equally likely ways to divide the N subsystems this way is

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

and so a similar line of argument results in

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

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

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.

<u>microcanonical ensemble</u> – 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 <u>unconstrained</u> 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

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

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

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

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

$\Rightarrow$ distribution that maximizes S is equally likely states

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

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

$\Rightarrow$ 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.

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

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

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

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

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

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

Determine β by condition that

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

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

More generally if we had any quantity X constrained, ie X_i is value in state i, and average value $\langle X \rangle = \sum_i p_i X_i$ is fixed, then

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

β 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.