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, i.e., the less (more) information we have about the probable state of the system.

We want \( S \) to satisfy the following properties:

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

2) For equally likely \( p_i \), i.e., all \( p_i = \frac{1}{N} \) 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 \( n = 1, \ldots, N \), and a second system with \( M \) equally likely states labeled by \( m = 1, \ldots, M \).

The combined system has \( N \times M \) equally likely 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 \( 1 - p_1 \). In general \( p_1 \neq p_2 \), i.e. the states need not be equally likely.
What is the disorder of the 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 the joint system of $N$ copies to be

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

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

$$P_M = \frac{N!}{M! (N-M)!} p_1^M p_2^{N-M} \quad \text{Prob} \left( M \text{ in state } 1 \right)$$

For $N$ large, this probability is very strongly peaked about the average $M = N p_1$. We have

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

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 $N p_1$ in state 1 and $N p_2$ in state 2.

How many ways are there to choose which of the $N$ two level sub-systems are in state 1?
There are \( \frac{N!}{(N_{p_1})! (N_{p_2})} \) ways, \( (N_{p_2} = N(1-p_2)) \)

each of these ways are equally likely!

\( \Rightarrow \) the entropy of the N copy system is

\[
S = k \ln \left[ \frac{N!}{(N_{p_1})! (N_{p_2})} \right] \log \text{ of # equally likely states!}
\]

\[
= k \left[ \ln N! - \ln (N_{p_1})! - \ln (N_{p_2})! \right]
\]

we Stirling formula

\[
= k \left[ N \ln N - N - N_{p_1} \ln N_{p_1} + N_{p_1} - N_{p_2} \ln N_{p_2} + N_{p_2} \right]
\]

use \( N_{p_1} + N_{p_2} = N \) as \( p_1, p_2 = 1 \)

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

\( \Rightarrow S = kN \left[ -p_1 \ln p_1 - p_2 \ln p_2 \right] \quad \text{since } p_1 + p_2 = 1 \)

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

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

Similarly, if our system had \( m \) possible states, with

probabilities \( p_1, p_2, \ldots, p_m \), and we took \( N \) copies of

this \( m \)-level system, the joint system would have

\( N_{p_1} \) subsystems in state 1, \( N_{p_2} \) in state 2, \ldots, \( N_{p_m} \) in state \( m \).

The number of equally likely ways to divide the \( N \) subsystems

this way is \( \frac{N!}{(N_{p_1})! (N_{p_2})! \cdots (N_{p_m})} \).
And so a similar line of argument results in

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

\[ S(\varepsilon, \pi) = -k \sum_i \pi_i \ln \pi_i \]

We define 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.
micromcanonical 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

\[ S + \lambda \sum_i p_i \]

Where \( \lambda \) is the Lagrange multiplier - we then determine the value of \( \lambda \) by imposing the constraint.

So if there are \( N \) states of energy \( E \), the maximization gives

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

\[ \Rightarrow p_i \left( \frac{1}{p_i} \right) + k \lambda = \lambda \]

\[ 1 + \ln p_i - \lambda = 0 \]

\[ p_i = e^{\lambda - 1} \] same for all \( i \)
A distribution that maximizes $S$ is equally likely states

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

$\Rightarrow$ in microcanonical ensemble at energy $E > 0$ 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.

$$
0 = \frac{2}{\partial P_i} \left( -k \sum_i \left[ P_i \ln P_i - \alpha P_i + \beta P_i E_i \right] \right)
$$

$\Rightarrow 0 = 1 + \ln P_i - \alpha + \beta E_i$

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

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

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

$$
\Rightarrow P_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}
$$
If we interpret $\beta = \frac{1}{k_B T}$, we recover the canonical distribution.

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

$$p_i = \frac{e^{-\beta X_i}}{\sum_j e^{-\beta X_j}}$$

gives maximum $S$ consistent with the constraint.

$\beta$ 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.
Grand Canonical Ensemble

Consider a system of interest which is in contact with both a thermal and a particle reservoir.

- System of interest $E, V, N$
- Wall allows exchange of energy and particles
- Reservoir $E_R, V_R, N_R$

One way such a situation may arise physically is if the "system of interest" is just a certain volume immersed in a much larger volume of the same "stuff", and the walls around the "system of interest" are just our mental constructs.

- Gas in a box
- Reservoir is the rest of the gas

System of interest
- is some interior region of the gas. Dashed lines are mental constructs - not physical walls!
- The energy $E$ and number of particles $N$ in the region of interest are not fixed but fluctuate as energy and particles flow between the region and the rest of the gas.
The reservoir is so large, that no matter how much energy or particles the system of interest transfers to it, its temperature $T_R$ and chemical potential $\mu_R$ do not change — this is what we mean by it being a reservoir.

We see this as we argued before. If heat $dQ = Tds$ is transferred to the reservoir then the change in $T_R$ is

$$\Delta T_R = \frac{\partial T_R}{\partial S_R} ds = \left(\frac{\partial^2 E_R}{\partial S_R^2}\right) ds \sim \frac{N}{N_R} \frac{T_R}{T_R} \text{as } E_R, S_R \sim N_R \text{ and } ds \sim N \text{ at most}$$

so if $N \ll N_R$, $\Delta T_R \ll T_R$

Similarly, if $dN$ is transferred to the reservoir

$$\Delta N_R = \frac{\partial N_R}{\partial N} dN = \left(\frac{\partial^2 E_R}{\partial N_R^2}\right) dN \sim \frac{N}{N_R} \frac{\mu_R}{N_R} \text{as } E_R, N_R \sim N_R \text{ and } dN \sim N \text{ at most}$$

so if $N \ll N_R$, $\Delta \mu_R \ll \mu_R$

So we regard $T_R$ and $\mu_R$ of the reservoir as fixed.

Now, because the system of interest is in equilibrium with the reservoir, we have $T = T_R$, and $\mu = \mu_R$. 
Now \( N + N_R = N_T \) is fixed
\( E + E_R = E_T \) is fixed
\( V, V_R \) are fixed

As we discussed in the case of the canonical distribution, the total number of states available to the combined system + reservoir will be

\[
\Omega_T(E_T, V, V_R, N_T) = \int \frac{dE}{\Delta} \sum_N \Omega(E, V, N) \Omega_R(E_T - E, V_R, N_T - N) \uparrow \uparrow \\
\text{# states in system} \quad \text{# states in reservoir}
\]

\[
= \int \frac{dE}{\Delta} \sum_N \Omega(E, V, N) \exp \left\{ \frac{S_R(E_T - E, V_R, N_T - N)}{k_B} \right\}
\]

Prob for system to have \( E \) and \( N \) is

\[ P(E, N) \propto \Omega(E, V, N) \exp \left\{ \frac{S_R(E_T - E, V_R, N_T - N)}{k_B} \right\} \]

\[ S_R(E_T - E, V_R, N_T - N) = S_R(E_T, V_R, N_T) + \left( \frac{\partial S_R}{\partial E_R} \right)(-E) + \left( \frac{\partial S_R}{\partial N_R} \right)(-N) \]

\[ = \text{const} - \frac{E}{T} + \frac{\mu N}{T} \]

\[ \Rightarrow P(E, N) \propto \Omega(E, V, N) \exp \left\{ -\frac{(E - \mu N)}{k_B T} \right\} \]

Normalizing \( P(E, N) = \frac{\Omega(E, V, N) \exp \left\{ -\frac{(E - \mu N)}{k_B T} \right\}}{\sum_N \frac{dE}{\Delta} \Omega(E, V, N) \exp \left\{ -\frac{E}{k_B T} + \frac{\mu N}{k_B T} \right\}} \)
\[ p(E, N) = \frac{\sum_{N} (E, V, N) \ e^{-(E - \mu N)/k_B T}}{\sum_{N} \Omega_N(V, T) \ z^N} \]

where \( z = e^{\mu/k_B T} \) is called the fugacity.

Define the grand canonical partition function

\[ \mathcal{Z}(z, V, T) = \sum_{N=0}^{\infty} z^N \Omega_N(V, T) \]

\[ = \sum_{N} \int_{\Delta} \frac{dE}{\Delta} \Omega(E, V, N) e^{-(E - \mu N)/k_B T} \]

More generally, if the states of the system are labeled by an index \( i \), and state \( i \) has energy \( E_i \) and particle number \( N_i \), then

\[ \mathcal{Z}(z, V, T) = \sum_{i} z \ e^{-(E_i - \mu N_i)/k_B T} \]

and

\[ P_i = \frac{e^{-(E_i - \mu N_i)/k_B T}}{\mathcal{Z}(z, V, T)} \]

Note: These expressions make no reference to the reservoir.
Alternatively - for classical indistinguishable particles

Consider system + reservoir to be a a fixed $T$ in a canonical ensemble

Canonical partition function for system + reservoir, with
volume $V_T = V + V_R$ and number particles $N_T = N + N_R$, is

$$Q_{N_T}(T, V_T) = \frac{1}{\hbar^{3N_T} N_T!} \prod_{i=1}^{3N_T} \left( \int_{V_T}^{} \int_{V_R}^{} d\mathbf{q}_i d\mathbf{p}_i \right) e^{-\beta H_T}$$

$H_T$ is total Hamiltonian.

Imagine dividing the combined system into the "system of interest" with $N$ particles in $V_T$, and the reservoir with $N_R$ particles in $V_R$.

The system of interest is weakly interacting with the reservoir, so

$$H_T = H + H_R$$

\[\text{system of interest} \; \overset{\rightarrow}{\text{reservoir}}\]

and

$$\int_{V_T} d\mathbf{q}_i = \int_{V_T}^{} d\mathbf{q}_i = \int_{V_R}^{} d\mathbf{q}_i$$

$$Q_{N_T}(T, V_T) = \frac{1}{\hbar^{3N_T} N_T!} \prod_{i=1}^{3N_T} \left( \int_{V}^{} d\mathbf{q}_i + \int_{V_R}^{} d\mathbf{q}_i \right) \int_{V_T}^{} d\mathbf{p}_i e^{-\beta H} e^{-\beta H_R}$$

Expand out the product of factors - each term will correspond to a certain number $N$ particles in $V_T$, and the remainder $N_R = N_T - N$ in $V_R$. 
Because the particles are indistinguishable, it does not matter which \( N \) of the \( N_T \) are in \( V \) and which \( N_R \) are in \( V_R \). Each such term contributes the same amount. We can therefore consider just one such term, and multiply it by the number of ways to put \( N \) in \( V \), with the remainder in \( V_R \).

The number of such ways is \( \frac{N_T!}{N! N_R!} \).

\[
Q_{N_T}(T, V_T) = \frac{1}{h^{3N}} \sum_{N=0}^{N_T} \frac{N_T!}{N! N_R!} \left( \prod_{i=1}^{3N} \int_{\mathbb{R}^3} d\mathbf{r}_i \int_{\mathbb{R}^3} d\mathbf{p}_i e^{-\beta H} \right) \left( \prod_{j=1}^{3N_R} \int_{\mathbb{R}^3} d\mathbf{r}_j \int_{\mathbb{R}^3} d\mathbf{p}_j e^{-\beta H_R} \right)
\]

\[
= \sum_{N=0}^{N_T} \left( \frac{1}{h^{3N}} \prod_{i=1}^{3N} \int_{\mathbb{R}^3} d\mathbf{r}_i \int_{\mathbb{R}^3} d\mathbf{p}_i e^{-\beta H} \right) \left( \frac{1}{h^{3N_R}} \prod_{j=1}^{3N_R} \int_{\mathbb{R}^3} d\mathbf{r}_j \int_{\mathbb{R}^3} d\mathbf{p}_j e^{-\beta H_R} \right)
\]

\[
Q_{N_T}(T, V_T) = \sum_{N=0}^{N_T} Q_N(T, V) Q_{N_R}(T, V_R)
\]

The probability that there are \( N \) particles in \( V \) is therefore proportional to the weight this term has in the above sum.

\[
\Phi(N) \propto Q_N(T, V) Q_{N_R}(T, V_R) = Q_N(T, V) e^{-A(T, V_R, N_R)/k_B}
\]

Expand

\[
A(T, V_R, N_R) = A(T, V_R, N_T - N)
\]

\[
= A(T, V_R, N_T) - \left( \frac{\partial A}{\partial N} \right)_{T, V_R} N
\]

\[
= \text{const} - \mu N
\]

\[\text{independent of } N\]
So

\[ P(N) \propto Q_N(T, V) e^{\mu N/k_B T} \]

\[ P(N) = \frac{Q_N(T, V) e^{\mu N/k_B T}}{\sum_{N=0}^{\infty} Q_N(T, V) e^{\mu N/k_B T}} \]

where we set \( N_T \to \infty \) in upper limit of sum.

Define \( Z = e^{\mu/k_B T} \)

Grand canonical partition function

\[ Z(z, T, V) = \sum_{N=0}^{\infty} Q_N(T, V) e^{\mu N/k_B T} \]

Substitute for \( Q_N \) to get

\[ P(N) = \frac{\int_{\Delta E} \Omega(E) e^{-E/k_B T} e^{\mu N/k_B T}}{Z} \]

or \[ P(E, N) = \frac{\Omega(E) e^{-(E-\mu N)/k_B T}}{Z} \]

as before.