Statistical Mechanics

General question:
What type of probabilities of microstates \( i \) should we assume if we only know the quantities \( M^5 \) and nothing else?

Principle idea of SM

Define entropy function ( = negative information measure ).
Maximize entropy subject to given constraints.
Then probabilities that maximize the entropy are the relevant ones that correctly describe the system under consideration.
Information measures

Consider events labelled by \( i \in N \).
Probabilities for these events to happen are \( P_i \).
\[
\sum_{i=1}^{W} P_i = 1
\]

Suppose event \( j \) occurs.
Suppose \( P_j = 1 \), whereas \( P_i = 0 \) \((i \neq j)\).

\[\Rightarrow\] we learn very little from the actual occurrence of the event.

But if \( P_i = \frac{1}{W} \) \((i)\) then we learn at least from the actual occurrence of event \( j \).

Information gain due to occurrence of an event is given by some function \( h(P_i) \).

For example, choose \( h(P_i) = \log P_i \) in a long sequence of independent
trials, the average information gain is given by

\[ I(\{p_i\}) = \sum_{i=1}^{n} p_i \cdot h(p_i) \]

The entropy \( S \) is then defined as

\[ S = -I \]

The "Khinchin Axioms"

These general axioms describe what properties an information measure should have.

Axiom 1

\[ I = I(p_1, p_2, \ldots, p_w) \]

Information should only depend on frequencies of events and nothing else.

Axiom 2

\[ I(\frac{1}{w}, \ldots, \frac{1}{w}) \leq I(p_1, \ldots, p_w) \]
The information measure \( I \) takes on a minimum for uniform distribution.

**Axiom 3**

\[
I(P_1, \ldots, P_n) = I(P_1, \ldots, P_n, 0)
\]

This means the information measure should not change if the sample set is enlarged by another event that has prob. 0.

**Axiom 4**

\[
I\{P_i, P^{\bar{i}}\} = I\{P_i\} + \sum\limits_j P_i I(P^{\bar{i}}(j|\bar{i}))
\]

**Meaning:** The way the information is collected should be independent of how we divide the entire system into subsystems.
In general, systems I and II are not statistically independent. But, in case they are, then
\[ P_{ij} = P_i \cdot P_j \]
and thus
\[ I(\{P_{ij}\}; P) = I(\{P_i\}; P) + I(\{P_j\}; P) \]

since \[ P(j|i) = P_i \]

**Important results**

The Shannon entropy
\[ S := -K \sum_{i=1}^{\infty} P_i \log P_i = -I \]

is the only solution to the Schrödinger axioms 1-4.

**Proof**

in one of the next lectures.

Often \( K = 1 \)
Other information measures

Rényi entropy

\[ S_q^{(R)} = \frac{1}{q-1} \ln \sum_i p_i^q \]

\( q \in \mathbb{R} \) a parameter

The summation is over all events with \( p_i \neq 0 \).

This satisfies versions 1-3 but not 4.

For \( q \to 1 \) \( S_q^{(R)} \) reduces to the Shannon entropy.

Proof

\( q = 1 + \varepsilon \) \( \varepsilon \) small

\[ S_{1+\varepsilon}^{(R)} = \frac{1}{\varepsilon} \ln \sum_i p_i^{1+\varepsilon} \]

\[ = \frac{1}{\varepsilon} \ln \sum_i p_i \cdot p_i^\varepsilon \]

\[ \approx e^{\varepsilon \log p_i} \equiv 1 + \varepsilon \log p_i \]
\[
\frac{-1}{e} \log \left( \sum_i \pi_i \left( 1 + e \log \pi_i \right) + \ldots \right) \\
= \frac{1}{e} \log \left( 1 + e \sum_i \pi_i \log \pi_i + \ldots \right) \\
= \frac{1}{e} \left( \not{\sum_i \pi_i \log \pi_i} + \ldots \right) \\
= \sum_i \pi_i \log \pi_i \\
\text{if } e \gg 0
\]

The Tsallis entropy

\[
S_q(T) = \frac{1}{q-1} \left( 1 - \sum_{i=1}^n \pi_i^q \right)
\]

Exercise 1

Show that for \( q \gg 1 \) the Tsalli entropy reduces to the Shannon entropy.

Among these types of entropies satisfy axioms 1-3 but not axiom 4.
Further entropies

Landsberg-Rednal entropy

\[ S^{(L)}_q = \frac{1}{q-1} \left( \frac{1}{\sum_{i=1}^{q} p_i^q} - 1 \right) \]

The entropy

\[ S_{9, Abe} = - \sum_i \frac{p_i^9 - p_i^{9-n}}{9 - 9^{-n}} \]

Jaynes' information theory gives a simple information-theoretic derivation of this.

Assume we know some mean values \( M^0, \delta = 1, \ldots, s \) of an complex system under consideration.

Suppose we measure information with some information measure

\[ I(\{ p_i \}) = T \delta \]
What probabilities \( P_i \) for the microstates \( i \) of the system should we assume?

One information should have a minimum given the constraints of the system. Because otherwise we would have unjustified prejudices on the system, which we should have entered in terms of additional constraints \( M^5 \) in the first place.

\[ \Rightarrow I[p] \] must have minimum
\[ \Rightarrow S[p] = -I[p] \] must have a maximum (under the given constraints \( M^5 \)).

Constraints are given in form of the following conditions:
\[
\sum_{i=1}^{w} \pi_i M_i^6 = M^6 \quad (\delta = 1, \ldots, 5)
\]

Example: mean energy of a system of particles:
\[
\sum_{i=1}^{w} \pi_i E_i = U
\]

mean energy

associated with microstate i

Normalization condition is also regarded as a constraint:
\[
\sum_{i=1}^{w} \pi_i = 1
\]

To find the distributions that maximize the entropy under the given constraints, we introduce a function \( \Psi[p] \):

\[
\Psi[p] := S[p] + \sum_{\delta} \beta_\delta \left( \sum_{i} \pi_i M_i^\delta \right) \quad \text{Lagrange multiplier}
\]
Wanted to find minimum of

\[ \psi[p] = \min \text{ or } I[p] \text{ given the constraints} \]

Minimum can be obtained by differentiating:

\[ \frac{\partial}{\partial p_i} \psi[p] = 0 \quad (i=1, \ldots, W) \]

This means:

\[ \frac{\partial}{\partial p_i} I[p] + \sum \beta \sigma_i M_i = 0 \]

Ordinary material:

\[ I[p] = \sum_i p_i \log p_i \]

\[ \frac{\partial}{\partial p_i} I[p] = \log p_i + 1 \]

\[ \log p_i + 1 + \beta E_i = 0 \quad (\star) \]

Remind: for the canonical ensemble one has \( M_i = E_i \)

\[ \frac{1}{N} \sum p_i = 1 \]

Energy of i-th microstate.
\[ \beta_1 \equiv \alpha \quad \beta_2 \equiv \beta \]

\[ \sigma = 1, 2 \]

meaning

\[ \psi(p) = \sum p_i \log p_i + \alpha \sum p_i + \beta \sum p_i E_i \]

Solving (*) for \( p_i \) gives

\[ p_i = \frac{1}{Z} e^{-\beta E_i} \]

\( Z \): partition function

\[ Z = \sum e^{-\beta E_i} = e^{1 + \alpha} \]

**Definition**

The free energy \( F \) of any complex system with partition function \( Z(\beta) \) is defined as

\[ F = -\frac{1}{\beta} \log Z(\beta) \]

**Theorem**

For the canonical ensemble one has