Jaynes (1957): stat. phys. as an inference problem using limited data (such as \( \bar{E} \))

Maximize \( S = -k_B \sum_i p_i \log p_i \) 

\( p_i \): prob. to be in cell \( i \) of phase space

With constraints:

1. \( \sum_i p_i = 1 \) normal 'n

2. \( \langle E \rangle = \sum_i p_i E_i = \bar{E} \)

Energy of a system of particles

\[ S \left[ \frac{\sum_i}{k_B} - \beta \left( \sum_i p_i E_i - \bar{E} \right) - \log \left( \sum_i p_i - 1 \right) \right] = 0 \]

wrt \( \{ p_i \} \) & \( \beta \)

\[ P_i^* = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}} \] \( \{ P_i^* \} = Q \), part'n f'n

\[ S_{\text{max}} = -k_B \sum_i P_i^* \log P_i^* = -k_B \left[ \sum_i P_i^* (-\beta E_i) - \sum_i P_i^* \left( \frac{\sum_j e^{-\beta E_j}}{Q} \right) \right] = \]

\[ + k_B \log Q + k_B \beta \langle E \rangle = \frac{\langle E \rangle - F}{1/T} \]

Where

\[ F = -k_B T \log Q \] is the free en.

Max of entropy $\Rightarrow$ self-consistent way to draw inferences from prob. distrb's.

Max of $H$ with constraints should be:

1) unique
2) invariant wrt coord. transform'n
3) relative prob. of two subsets of outcomes should not depend on other subsets if data are provided indep. for each subset
4) prob. for $2$ indep. systems should be a product of marginal probs.

Consider prob's: prior prob's.

$$H(p_1, p_2) = - \sum_i (p_i \log p_i - \bar{a})$$

--- some property of interest ---

$p_i \rightarrow q_i$ when there are no constraints (i.e., no data)

Constraints lead to $p_i \neq q_i \Rightarrow$ discard $\bar{a}$

Start with

$$(\Delta p_i, \Delta p_k) [H - \sum_i (p_i \log p_i)]$$

E.g. $p_j$ increased, $p_k$ decreased $\Rightarrow$

$\Rightarrow$ other cells $e_{ij, k}$ unaffected

(axiom 3):
\[ p_e(\partial p_j - \partial p_k)(H - \lambda \sum_i \partial \phi_i \partial p_i) = 0 \]
\[ l \neq k, i \]
\[ (\partial p_j - \partial p_k)(H) \text{ must depend only on } j, k \]
\[ \exists p_m H \text{ must depend only on } m \]
\[ H = \sum f(p_i, q_i) \]

"cell-level" quantities

Continuous representation:

\[ H = \int Dx \, f(p(x), q(x)) \]

Coord. inv. (axiom 2):

\[ x \rightarrow y \Rightarrow Dy = Dx \cdot J \]

\[ \int Dx \, p(x) = 1 \Rightarrow \int Dy \frac{p(y)}{Dx \cdot J} = 1 \]

\[ \int Dx \, p(x) a(x) = \int Dy \frac{p(y) a(y)}{Dx \cdot J} = a \]

\[ a(x) = a(y) \]

So,

\[ H' = \int Dx \, J f \left( \frac{p(x)}{J}, \frac{q(x)}{J} \right) \]
\[ \delta p(x) \left[ H - \lambda \int Dx \, p(x) a(x) \right] = 0 \text{ gives} \]

\[ \frac{\partial}{\partial p(x)} \left[ \mathcal{L}[p(x), q(x)] \right] - \lambda a(x) = 0. \]

\[ \Rightarrow \quad \mathcal{L}[p(x), q(x)] \]

\[ \frac{\delta}{\delta p(y)} \left[ H' - \lambda' \int Dy \, p(y) a(y) \right] = \]

\[ = \mathcal{L} \left[ \frac{p(x)}{J}, \frac{q(x)}{J} \right] - \lambda' a(y) = 0 \]

So,

\[ \mathcal{L} \left[ \frac{p(x)}{J}, \frac{q(x)}{J} \right] = (\lambda - \lambda) a(x) + \mathcal{L}[p(x), q(x)] \]

Since \( J \) is arbitrary, it has to vanish from LHS:

\[ \mathcal{L} \left[ \frac{p(x)}{J}, \frac{q(x)}{J} \right] = \mathcal{L} \left[ \frac{p(x)}{q(x)} \right] \]

\[ \Rightarrow \quad \mathcal{L}[p(x), q(x)] = p(x) \mathcal{L} \left[ \frac{p(x)}{q(x)} \right] + J \mathcal{L}[q(x)] \]

Indeed,

\[ \frac{\partial}{\partial p(x)} \left[ \mathcal{L}[p(x), q(x)] \right] = h \left[ \frac{p(x)}{q(x)} \right] + h' \left[ \frac{p(x)}{q(x)} \right] \frac{p(x)}{q(x)} \]

\[ h \left[ \frac{p(x)}{q(x)} \right] \text{ is arbitrary thus far} \]
Finally, **axiom 4 (system indep.)**: consider two indep. systems described by \( x_1 \& x_2 \):

\[ \int D(x) A_k(x_k) p(x_1, x_2) = A_k \quad k=1,2 \]

Define \( R(x) = \frac{p(x)}{g(x)} \), \( x = \{x_1, x_2\} \)

Now, \( H = \int D(x) p(x) h(R(x)) \) gives

\[
\frac{\delta}{\delta p(x)} \left[ H - \lambda_1 \int D(x) p(x_1, x_2) a_1(x_1) - \lambda_2 \int D(x) p(x_1, x_2) a_2(x_2) \right] =
\]

\[
= h(R(x)) + r(x) h'(R(x)) - \lambda_1 a_1(x_1) - \lambda_2 a_2(x_2) = h[r_1 r_2] + r_1 r_2 h'[r_1 r_2] - \lambda_1 a_1 - \lambda_2 a_2 = 0.
\]

\[
\frac{\partial^2}{\partial x_1 \partial x_2} : \frac{\partial}{\partial x_1} \left\{ h'[r_1 r_2] (\partial r_1 / \partial r_1 r_1') + r_1 r_2 h''[r_1 r_2] (\partial^2 r_1 / \partial r_1^2 r_1') + r_1 r_2 r_1 h'[r_1 r_2] - \right. \]

\[
- \lambda_2 a_2 ' \right\} =
\]

\[
= h''[r_1 r_2] (r_1' r_2) (r_2' r_1') + h'[r_1 r_2] (r_2' r_1') + h'''[r_1 r_2] (r_1' r_2) (r_2' r_1') +
\]

\[
+ h''[r_1 r_2] (2r_1 r_1' r_2 r_2') + r_1' r_2' h'[r_1 r_2] +
\]

\[
+ r_1 r_2 h'[r_1 r_2] (r_1' r_2) =
\]

\[
r_1 r_2 \left\{ 4 r_1 r_2 h'' + 2 h' + r_1^2 r_2^2 h''' \right\} = 0 \quad \text{or}
\]

\[
(5)
\]
\[ y^2 h'' + 2h' + r^2 h'' = 0 \]

\[ \text{"hir2 (system indep.)"} \]

\[ h(r) = -K \log(r) + B + C/r \]

\[ \text{scale factor, } >0 \quad \text{\#some const} \]

\[ \text{(just a const in } H) \]

\[ \text{So, } H = \int \mathcal{D}[x] p(x) \left[ -K \log(p_x) + B + C \frac{q_x}{p(x)} \right] = \]

\[ = -K \int \mathcal{D}[x] p(x) \log \left( \frac{p(x)}{q(x)} \right) + \]

\[ + C \int \mathcal{D}[x] q_x(x) \]

\[ \text{also an irrelevant const} \]

\[ \text{Note that} \]

\[ H = -K \int \mathcal{D}[x] p_1 p_2 \left[ \log \frac{p_1}{q_1} + \log \frac{p_2}{q_2} \right] = \]

\[ = -K \left[ \int \mathcal{D}x_1 p_1 \log \frac{p_1}{q_1} + \int \mathcal{D}x_2 p_2 \log \frac{p_2}{q_2} \right] = \]

\[ = H_1 + H_2, \text{ as expected} \]

\[ \text{Note that } H \text{ is a KL distance between } p(x) \text{ and } q(x) \]

\[ \text{Maximize } H \text{ to minimize distance between } p(x) \text{ and } q(x) \text{ given the constraints.} \]
Types of constraints
higher moments, combinations of moments?
shall we use $\int \int p(x) a(x)$
Nonlinear in $p(x)$ constraints?
Axiom 1: unique max of $H$ under constraints $\Rightarrow$ no constraints that would mess up the convexity of
$\int \int p(x) \log p(x)$
What should $a(x)$ be? Could be higher moments, but not always useful...

Thermodynamics: a system is in contact with a large bath

$\text{system + bath} = \overset{\text{closed}}{\otimes}$

Then $F = - \sum_i p_i \log p_i + V(\sum_i p_i a_i - 1) + \lambda \sum_i p_i a_i (1 - \sum E_i + E_a, E_{tot})_{\text{const}}$

$p_{i,a} = 0 \& E_i + E_a \neq E_{tot}$

Note that $p_i = \sum_a p_{i,a}$,
$p(\alpha i) = \frac{p_{i,a}}{p_i} \\longleftrightarrow p_i = p(\alpha i) p_i$ [indep. not assumed]

$\sum_i p_i = 1 \, , \, \sum_a p(\alpha i) = 1$

Then $- \sum_i p_i \log p_i = \sum_{i,a} p(\alpha i) p_i \left[ \log p(\alpha i) + \log p_i \right] = - \sum_i p_i \log p_i - \sum_{i,a} p(\alpha i) p_i \log p(\alpha i)$.
So, 
\[ F_{\text{new}} = -\sum_i p_i \log p_i - \sum_{i,a} p(a_{li}) p_i \log p(a_{li}) + \]
\[ + \sum_i V_i \left[ \sum_a p(a_{li}) - 1 \right] \delta_{E_i + E_a, E_{\text{tot}}} + \lambda \sum_{i,a} p(a_{li}) p_i \left[ 1 - \delta_{E_i + E_a, E_{\text{tot}}} \right]. \]

\[ \delta F_{\text{new}} \text{ w.r.t. } p(a_{li}), V_i, \lambda : \]
\[ \sum_{i,a} p(a_{li}) p_i \left[ 1 - \delta_{E_i + E_a, E_{\text{tot}}} \right] = 0, \]
\[ \sum_a p(a_{li}) = 1, \text{ if } E_i + E_a = E_{\text{tot}} \]
\[ -p_i \log p(a_{li}) - p_i + V_i + \lambda \pi_i \left[ 1 - \delta_{E_i + E_a, E_{\text{tot}}} \right] = 0 \]
\[ \log p(a_{li}) = \begin{cases} \frac{V_i}{p_i} - 1, & E_i + E_a = E_{\text{tot}} \\ \frac{V_i}{p_i} + \lambda - 1, & E_i + E_a \neq E_{\text{tot}} \end{cases} \]

\[ \sum_{i,a} p(a_{li}) p_i = 0, \text{ or } p(a_{li}) = 0, \quad i,a \]

\[ \sum_i \frac{V_i}{p_i} + \lambda - 1 \pi_i = 0 \Rightarrow \left( \sum_i \frac{V_i}{p_i} \right) = 0 \]

\[ \sum_{i,a} \frac{V_i}{p_i} + \lambda - 1 = 1 \Rightarrow \left( \sum_i \right)_{\pi_i} = \infty \]

\[ \sum_i \frac{V_i}{p_i} + \lambda = 0 \] consistent with \( \lambda = -\infty \) 

\[ \sum_{i,a} \frac{V_i}{p_i} p_i = 0 \Rightarrow \left( \sum_{i,a} \frac{V_i}{p_i} \right) = 0 \]

\[ \sum_{a} \frac{V_i}{p_i} + \lambda - 1 = 0 \Rightarrow \left( \sum_{a} \frac{V_i}{p_i} \right) = -\infty \]

\[ \text{for microstates at } E_a \neq E_{\text{tot}}, E_i = \emptyset \]

\[ E_i + E_a = E_{\text{tot}}: \]
\[ \sum_{a} \frac{V_i}{p_i} + \lambda - 1 = 0 \Rightarrow p(a_{li}) = \frac{1}{\Omega (E_{\text{tot}} - E_i)} \]

\[ \sum_{i,a} \frac{V_i}{p_i} - 1 = 1 \Rightarrow \text{indep. of } a \]

\[ \text{So, } p(a_{li}) = \frac{\delta_{E_a, E_{\text{tot}} - E_i}}{\Omega (E_{\text{tot}} - E_i)} \]

8
Now,
\[ F_{new} = - \sum_i p_i \log p_i + \Delta \left( \sum_i p_i - 1 \right) - \]
\[ - \sum_{i,a} p_i \frac{\delta E_a, E_{tot} - E_i}{\mathcal{Z}(E_{tot} - E_i)} \log \frac{\delta E_a, E_{tot} - E_i}{\mathcal{Z}(E_{tot} - E_i)} = \]
\[ = - \sum_i p_i \log p_i + \Delta \left( \sum_i p_i - 1 \right) + \sum_i p_i \log \mathcal{Z}(E_{tot} - E_i) \]

Large bath: \( E_{tot} \gg E_i \),
\[ \log \mathcal{Z}(E_{tot} - E_i) \approx \log \mathcal{Z}(E_{tot}) - \beta E_i \]

so the constraint becomes
\[ -\beta \sum_i p_i E_i \]
just as before.

Higher-order moments drop out when the bath is large \( \rightarrow 1\text{st order moments arise naturally.} \)

Indeed, as \( N \to \infty \) in any system,
\[ Q^{\text{basic}} = \sum_E g(E) E^{-\beta E} \]
\( \text{energy level degeneracy} \)

fluct's \( \langle E^2 \rangle - \langle E \rangle^2 \) become
\( \text{small as } N \uparrow \),
\( (\text{as } \frac{1}{N}) \)
only \( \langle E \rangle \) remains finite

Higher-order moments also vanish...

In "nanothermodynamics", higher-order moments may play a role, but the MAXENT framework still holds.