1.2.1) The microcanonical ensemble

Microcanonical hypothesis For a sufficiently complex system, the energy surface is visited uniformly & ergodically $\Rightarrow$ All configurations with the same energy are visited with equal probability.

Microcanonical measure for discrete systems

Classical isolated system described by a set of configurations $\{\varphi\}$. Then, if the system is at energy E, its microcanonical distribution is

$$\boxed{P_E(\varphi) = \frac{1}{\Omega(E)} \delta_{H(\varphi),E}} \qquad (1)$$

where: * $\Omega(E)$ is the number of configurations of energy E

* $\delta_{a,b}$ is the KROENECKER delta, such that $\delta_{a,b} = 1$ if $a=b$ & $\delta_{a,b} = 0$ otherwise

Continuous systems There are mathematical subtleties to generalise (1) to continuous space that we will see in chapter 3. Essentially, P_E becomes a probability density and $\Omega(E)$ is the area of the energy surface of energy E.

Microcanonical entropy

. $\Omega(E)$ is a normalization constant such that $\sum_{\varphi} P_E(\varphi) = 1$

. $\Omega(E)$ varies with N & E, typically exponentially, so that it is better measured using Boltzmann microcanonical entropy

$$\boxed{S_m(E) = k_B \ln \Omega(E)} \qquad \text{with } k_B = 1.380\,649\ 10^{-23}\ J.K^{-1}$$

Microcanonical temperature

The variations of S_m & Ω with E are quantified by the microcanonical temperature

$$\boxed{\frac{1}{T_m} = \frac{\partial S_m}{\partial E}}$$

Comments:

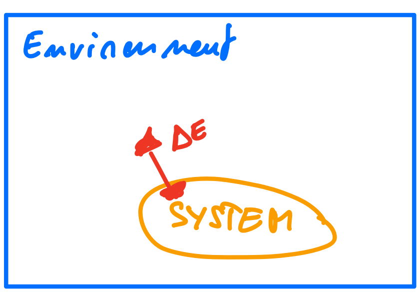
Q_1: Is (1) simple? Yes! As simple as it gets!

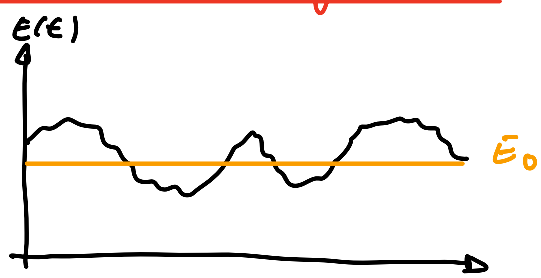
Q_2: Is (1) practical? Not really, computing $\Omega(E)$ is a combinatorics challenge

Q_3: Is (1) useful? Yes & No. We can engineer isolated systems (ultra high vacuum), but most systems are not isolated $\Rightarrow$ need to account for energy exchange & fluctuations

$\Rightarrow$ How to generalize (1) to open systems? Use information theory.

1.2.2) The principle of minimal information





The environment imposes fluctuations around $\langle E(q) \rangle = E_0$.

What $P(q)$ should we use? Two constraints:

① $\sum_q P(q) = 1$ & ② $\sum_q E(q)\, P(q) = E_0$

JAYNES (1957): We should use $P(q)$ that satisfies ① & ② and has no other biases $\Rightarrow$ Use information theory.

Shannon information theory (1948): Consider a distribution P that characterizes the result of sampling a random variable $n \in \{1, \dots, N\}$. ⑦

Q: Can we characterize the surprise $s(p(n))$ one should feel at observing a given sample n?

a) If $p(n) = 1$; $s(p(n)) = 0$ since there is no surprise

b) $s(p(n))$ decreases as $p(n)$ increases

c) the surprise of independent events should add up.

If $p(n_1, n_2) = p(n_1)\, p(n_2)$ then
$$\left.\begin{aligned} s(p(n_1, n_2)) &= s(p(n_1)) + s(p(n_2)) \\ &= s(p(n_1) \cdot p(n_2)) \end{aligned}\right\} s(ab) = s(a)s(b)$$

Shannon: a+b+c $\Rightarrow$ $s(p(n)) = -k \ln(p(n))$ with $k > 0$

Shannon entropy The average surprise of a distribution is called Shannon entropy

$$\boxed{S_S = \langle s(p(q)) \rangle = -\sum_q p(q) \ln p(q)}$$

Gibbs entropy (1906) Gibbs proposed a definition of entropy different from that of Boltzmann

$$\boxed{S_G = -k_B \sum_q p(q) \ln p(q)}$$

For the microcanonical ensemble,

$$S_G(E) = -k_B \sum_q \frac{1}{\Omega(E)} \delta_{E(q),E} \ln\left(\frac{1}{\Omega(E)} \delta_{E(q),E}\right)$$

$$= k_B \underbrace{\sum_{q|E(q)=E} \frac{1}{\Omega(E)}}_{=1} \ln \Omega(E) = k_B \ln \Omega(E)$$

$\Rightarrow$ Boltzmann & Gibbs (Shannon) entropy coincide exactly in the microcanonical ensemble.

1.2.3) The canonical ensemble

Jaynes idea is that, to minimize the biases of $p(q)$, we should maximize its surprise, constrained to satisfying $\sum_q p(q) = 1$ & $\sum_q E(q)\, p(q) = E_0$

Use Lagrange multiplyers & define:

$$\mathcal{L} = -\sum_q p(q) \ln p(q) + \alpha \left[1 - \sum_q p(q)\right] + \beta \left[E_0 - \sum_q p(q) E(q)\right]$$

* let's maximize $\mathcal{L}$ with respect to $p(q)$

$$\frac{\partial \mathcal{L}}{\partial p(q_i)} = -\ln p(q_i) - 1 - \alpha - \beta E(q_i) = 0 \Rightarrow p(q_i) = e^{-1-\alpha} e^{-\beta E(q_i)}$$

* Normalization fixes α through $e^{1+\alpha} = \boxed{\sum_q e^{-\beta E(q_i)} \equiv Z}$, where Z is called the partition function.
* β is then fixed by requiring $\langle E \rangle = E_0 = \frac{1}{Z} \sum_q E(q) e^{-\beta E(q)} = -\frac{1}{Z} \partial_\beta Z$

$$\Leftrightarrow \quad \boxed{E_0 = -\partial_\beta \ln Z(\beta)}$$

All this shows that the canonical distribution $\boxed{p(q) = \frac{1}{Z} e^{-\beta E(q)}}$ is the least biased distribution such that $\langle E(q) \rangle = E_0$, with β determined through $E_0 = -\partial_\beta \ln Z(\beta)$.

Comments:

(i) This can be generalized to other constraints (see Pset 1)

(ii) This is nice but our ignorance does not determine the laws of nature

(iii) Sam Edwards generalized this to granular media ⇒ does not work perfectly. If you try to apply this to active matter (e.g. bacterial suspensions), this fails horribly.

⇒ need alternate derivations

1.2.4) Conserved quantities & statistical independence of macroscopic volumes

cf. Landau & Lifschitz, Statistical mechanics.

Liouville's equation

Hamiltonian system comprising N particles with positions & momenta

$$\vec{q}_i = \begin{pmatrix} q_{i,x} \\ q_{i,y} \\ q_{i,z} \end{pmatrix} \quad \& \quad \vec{p}_i = \begin{pmatrix} p_{i,x} \\ p_{i,y} \\ p_{i,z} \end{pmatrix}$$ whose dynamics read

$$\dot{\vec{q}}_i = \frac{\partial H}{\partial \vec{p}_i} \quad \& \quad \dot{\vec{p}}_i = -\frac{\partial H}{\partial \vec{q}_i} \; ; \quad \frac{\partial}{\partial \vec{p}_i} = \begin{pmatrix} \frac{\partial}{\partial p_{i,x}} \\ \frac{\partial}{\partial p_{i,y}} \\ \frac{\partial}{\partial p_{i,z}} \end{pmatrix} \quad \& \quad \frac{\partial}{\partial \vec{q}_i} = \begin{pmatrix} \frac{\partial}{\partial q_{i,x}} \\ \frac{\partial}{\partial q_{i,y}} \\ \frac{\partial}{\partial q_{i,z}} \end{pmatrix}$$

The initial condition is sampled from some distribution

$$\rho_0(\{\vec{q}_i, \vec{p}_i\}) = \rho(\{\vec{q}_i, \vec{p}_i\}, t=0) \Rightarrow \rho(\{\vec{q}_i, \vec{p}_i\}, t) = \,?$$

10

Since $\vec{q}_i$ & $\vec{p}_i$ are advected by the flow $\dot{\vec{q}}_i(t)$, $\dot{\vec{p}}_i(t)$, we can define the (probability) current $\vec{j}(\{\vec{q}_i,\vec{p}_i\},t) = \rho(\{\vec{q}_i,\vec{p}_i\},t) \begin{pmatrix} \dot{q}_{1,x} \\ \dot{q}_{1,y} \\ \dot{q}_{1,z} \\ \vdots \\ \dot{p}_{N,x} \\ \dot{p}_{N,y} \\ \dot{p}_{N,z} \end{pmatrix}$ such that

the variations of ρ in any volume V is due to the flux of $\vec{j}$ through the area ∂V: $\dfrac{d}{dt}\int_V d\Gamma\, \rho(\{\vec{q}_i,\vec{p}_i\},t) = -\int_{\partial V} \vec{j}(\{\vec{q}_i,\vec{p}_i\},t)\cdot d\vec{S}$ with $d\Gamma = \prod_i d\vec{q}_i\, d\vec{p}_i$

$$\Leftrightarrow \int_V d\Gamma\, \partial_t \rho(\{\vec{q}_i,\vec{p}_i\},t) = -\int_V \vec{\nabla}\cdot\vec{j}(\{\vec{q}_i,\vec{p}_i\},t) \quad \text{(divergence theorem)}$$

$$\Leftrightarrow \partial_t \rho = -\vec{\nabla}\cdot\vec{j} \quad \text{since this holds for all volume } V.$$

$$\Leftrightarrow \partial_t \rho = -\sum_{i=1}^{N}\sum_{\alpha=x,y,z}\left\{\frac{\partial}{\partial q_{i,\alpha}}\left(\dot{q}_{i,\alpha}\,\rho\right) + \frac{\partial}{\partial p_{i,\alpha}}\left(\dot{p}_{i,\alpha}\,\rho\right)\right\}$$

$$= -\sum_{i=1}^{N} \frac{\partial}{\partial \vec{q}_i}\cdot\left(\frac{\partial H}{\partial \vec{p}_i}\rho\right) - \frac{\partial}{\partial \vec{p}_i}\cdot\left(\frac{\partial H}{\partial \vec{q}_i}\cdot\rho\right)$$

$$\Leftrightarrow \boxed{\partial_t \rho(\{\vec{q}_i,\vec{p}_i\},t) = -\sum_{i=1}^{N} \frac{\partial H}{\partial \vec{p}_i}\cdot\frac{\partial \rho}{\partial \vec{q}_i} - \frac{\partial H}{\partial \vec{q}_i}\cdot\frac{\partial \rho}{\partial \vec{p}_i} = -\{\rho, H\}}$$

where we have introduced the Poisson Bracket

$$\{A,B\} = \sum_{i=1}^{N} \frac{\partial A}{\partial \vec{q}_i}\cdot\frac{\partial B}{\partial \vec{p}_i} - \frac{\partial A}{\partial \vec{p}_i}\cdot\frac{\partial B}{\partial \vec{q}_i}$$

Liouville's theorem : $\rho(\{\vec{q}_i(t), \vec{p}_i(t)\}, t)$ is constant along a trajectory.

Chain rule:

$$\frac{d}{dt}\rho(\{\vec{q}_i(t), \vec{p}_i(t)\}, t) = \partial_t \rho + \sum_i \dot{\vec{q}}_i \cdot \frac{\partial \rho}{\partial \vec{q}_i} + \dot{\vec{p}}_i \cdot \frac{\partial \rho}{\partial \vec{p}_i} = -\{\rho, H\} + \{\rho, H\} = 0$$

Landau's trick:

(I) $\ln \rho$ is additive Two subsystem $S_1 = \{\vec{q}_1, \vec{p}_1\}$ & $S_2 = \{\vec{q}_2, \vec{p}_2\}$ that

are very large have weak correlations $\Rightarrow \rho_{1 \cup 2}(\{\vec{q}_1, \vec{q}_2, \vec{p}_1, \vec{p}_2\}) \simeq \rho_1(\{\vec{q}_1, \vec{p}_1\})\rho_2(\{\vec{q}_2, \vec{p}_2\})$

$$\Rightarrow \ln \rho_{1\cup 2} = \ln \rho_1 + \ln \rho_2$$

(II) The energy is the "only" other additive constant of motion

$\Rightarrow \ln \rho \propto E$, let's call $-\beta$ the proportionality constant

$\Rightarrow$ Boltzmann weight (actually more complicated $\Rightarrow \rho = \frac{1}{Z} e^{-\beta E}$)

This is a second, appealing justification of the Boltzmann weight for subsystems. Still not a derivation $\Rightarrow$ kinetic theory of gases.