Mechanics - Lecture 11, 4/17/2019

Fresh start: remainder of semester will introduce some central concepts of statistical mechanics. My main sources: chapter 5 of Chorin + Hald and chapter 3 of Bühler. However these are rather terse + lacking in examples; for a treatment you can sit down + read, with plenty of examples, I recommend Mark Tuckerman's book "Statistical Mechanics: Theory + Molecular Simulation"

Orientation: goal of stat mech is to consider "typical behavior," for systems too complex to describe in detail (answering questions involving macroscopic quantities). We'll do this by describing things probabilistically, though no explicit source of randomness is assumed.

Why is this reasonable?

a) Recall Poincaré's Recurrence Thm, proved as consequence of vol-preserving property of Hamiltonian flow in phase space. Vol pres flow can easily have very complicated orbits - even dense ones (a key example is irrational flow on a torus) - though simpler behavior is also possible (eg rational flow on a torus is periodic; so is that of a 2D system such as a pendulum)

b) Birkhoff's ergodic Thm - true once again for a vol-preserving flow $\varphi_t(x)$ with a finite-volume invariant set D - says that if D is "indecomposable" (not decomposable into 2 invariant regions, each of positive measure) then

$$\underbrace{\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\varphi_t(x))\, dt}_{\text{time average}} = \underbrace{\frac{1}{|D|}\int_D f\, dx}_{\text{spatial average}}$$

and "spatial averaging" amounts to expected value wrt a very special prob distrn (the uniform one).

Two issues:

1) the Hamiltonian is conserved (for Hamiltonian dynamics) so no region of phase space can be both invariant + indecomposable.
2) Are orbits dense in surface $H = \text{const}$? Hard to know, in most cases.

Ans to (1) is easy: we can focus on infinitesimal shell $\{x : H(x) \in [E, E+dE]\}$.

Ans to (2) is hard; usually we just assume it is so, unless there's reason to think otherwise (eg due to an additional conservation law).

We'll see soon (by considering a small system coupled to a large one) that uniform measure is not the only interesting one; the measure with weight

$$\rho = \text{const} \cdot e^{-\beta H}$$

is also of special interest (it defines the "canonical distribution").

A minimal requirement for a prob distrn to be suitable in this context is that it be <u>invariant</u> under the flow.

How to assess this? Well, in fluid dynamics an initial density $\rho^{\circ}(x)$ evolves under flow with velocity $u = \frac{d}{dt}\Big|_{t=0} \phi_t(x)$ by

$$\partial_t \rho + \mathrm{div}(\rho u) = 0, \qquad \rho\big|_{t=0} = \rho_0$$

[Proof: for any region D, if

$$\frac{d}{dt}\int_D \rho(t,x)\,dx = -\int_{\partial D} \rho\, u\cdot n$$

Then

$$\int_D \partial_t \rho + \mathrm{div}(\rho u)\; dx = 0\;;$$

true for any $D \Rightarrow \partial_t\rho + \mathrm{div}(\rho u) = 0$.
Note: we have assumed in this argument that there is transport only, with no diffusion.]

So: ρ is invariant iff $\mathrm{div}(\rho u) = 0$. When u is vol-preserving ($\mathrm{div}\, u = 0$) this is equiv to $u\cdot\nabla\rho = 0$.

Now specialize this to Hamiltonian flow on phase space $(q_1,\dots,q_N,p_1,\dots,p_N)$ with $\vec{u} = \left(\frac{\partial H}{\partial p_1},\dots,\frac{\partial H}{\partial p_N},\frac{-\partial H}{\partial q_1},\dots,\frac{-\partial H}{\partial q_N}\right)$: invariance says

$$\sum \frac{\partial H}{\partial p_j}\frac{\partial \rho}{\partial q_j} - \sum \frac{\partial H}{\partial q_j}\frac{\partial \rho}{\partial p_j} = 0$$

("Poisson bracket of H and ρ must be 0").
A sufficient condn for this is that

ρ be a function of H, say $\rho = G(H(q,p))$, since then

$$u \cdot \nabla \rho = G'(H)\; u \cdot \nabla H = 0$$

for the Hamiltonian flow $\vec{u}$.

As already indicated, the two key examples of greatest interest are

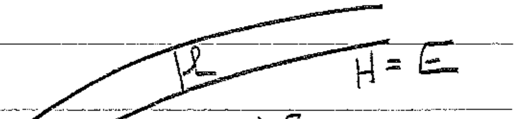
a) $\rho = \begin{cases} \text{const} & \text{for } H(q,p) \in [E, E+dE] \\ 0 & \text{otherwise} \end{cases}$

in limit $dE \to 0$ (this is the "<u>microcanonical ensemble</u>"

(b) $\rho = \text{const}\; e^{-\beta H(q,p)}$

(this is the <u>canonical ensemble</u>; as we'll see soon, β is essentially $1/\text{temp.}$)

Note about (a): this ρ is concentrated on surface $H = E$, but it is <u>not</u> a const times surface area. Rather, it is const times (surface area)$/|\nabla H|$, since (by "method of shells")



$$\int_{E<H<E+\Delta E} f\, d\text{vol} \simeq \Delta E \int_{H=E} \frac{f}{|\nabla H|}\, d\text{area} \qquad \ell \simeq \frac{\Delta E}{|\nabla H|}$$

Hypoth of (a) is that all pts in shell $\{E < H < E+dE\}$ are equally likely

The microcanonical distrn can be useful, but for most practical purposes the canonical one is more relevant.

1st pass: a quick (but not very physical) way to see its importance uses the information-theoretic notion of the entropy of a prob distrn. (In Bubber this is §5.6; in Chaikin + Hald it's §5.3).

Working with discrete prob distrns, for simplicity, let's discuss prob distrns $p_1, \dots, p_n$ on a finite set $\{1, \dots, n\}$. The (information theoretic) entropy is

$$S = -\sum p_i \ln p_i$$

(cont's analogue would be $-\int \rho \ln \rho \, dq \, dp$ where $\rho = \rho(q,p)$ is a prob density on phase space).

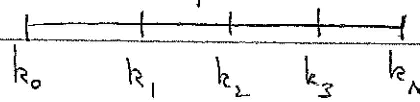
Question: what's so special about $p \ln p$? Ans (due to Shannon, I think):

with n states, $S = S(p_1, \ldots, p_n) = -\sum p_i \ln p_i$ has these properties:

(1) for each n, it is a cont's fn of n vars

(2) Let $S_n = S(1/n, \ldots, 1/n)$ be the entropy assoc to the uniform distrn on n pts. Then S_n is monotonically increasing in n.

(3) If we partition $(1, \ldots, n)$ into M bins



$1 = k_0 \le k_1 \le \cdots \le k_M = n$.

+ set $q_1 = p_1 + \cdots + p_{k_1}$, $q_2 = p_{k_1+1} + \cdots + p_{k_2}$, etc.

Then

$$S(p_1, \ldots, p_n) = S(q_1, \ldots, q_M) + q_1 S\left(\frac{p_1}{q_1}, \ldots, \frac{p_{k_1}}{q_1}\right) + q_2 S\left(\frac{p_{k_1+1}}{q_2}, \ldots, \frac{p_{k_2}}{q_2}\right) + \cdots$$

Moreover, S is the only such fn (up to a mult. constant). Interptn: "entropy" represents "uncertainty":

(1) $\iff$ entropy depends cont'sly on probabilities

(2) $\iff$ more states, each equally likely $\Rightarrow$ more uncertainty

(3) $\iff$ uncertainty is additive; sum of that inherent in a particular grouping + averages of uncertainties of the groupings.

Note that $p \to -p\ln p$ is concave, so

$$\max_{\substack{\sum' p_j = 1 \\ p_j \geq 0}} -\sum' p_j \ln p_j$$

is a concave maximization. The optimum is achieved when $p_1 = \cdots = p_n = 1/n$, by method of Lagr mult, since

$$\frac{\partial}{\partial p_j}\left[\sum' p_j \ln p_j - \lambda \sum' p_j\right] = 0 \iff 1 + \ln p_j = \lambda \text{ for each } j$$

No surprise – a localized prob. density carries lots of info, a uniform one carries none.

More interesting: canonical distrn arises by maximizing entropy subject to constraint on avg H

$$\max_{\substack{\sum' p_j = 1 \\ \sum' p_j H_j = U}} -\sum' p_j \ln p_j \quad \Rightarrow \quad p_j = Z e^{-\beta H_j} \text{ for some } \beta$$

Pf is again method of Lagr mult:

$$\frac{\partial}{\partial p_j}\left[\sum' p_j \ln p_j - \lambda \sum' p_j - \mu \sum' p_j H_j\right] = 0$$

$$\Rightarrow \ln p_j = \lambda + \mu H_j - 1$$

$$\Rightarrow \quad p_j = Z e^{-\beta H_j}$$ renaming the constants

Thus: canonical distn is the max-entropy distn consistent with a given value of $E[H]$.

Well, that's nice, but it isn't very physical (why should $E[H]$ be specified? why should entropy be maximized?) Answer is: we should use canonical distn when we're observing a (possibly small) system that's in equilibrium with a much larger one (a "reservoir"). To explain, I'll follow Bühler §3.2-3.3

Discn is fundamentally about probability, not dynamics, so we'll focus on a discrete "coin-flipping" example involving no dynamics:

$$\text{state} = X = (s_1, \ldots, s_N), \quad \text{each } s_i = \pm 1$$

$$\text{Hamiltonian} = H(X) = \sum_{i=1}^{N} s_i$$

(if s_i reports coin flip by $+1$ = heads, -1 = tails then H reports the # of heads in N flips).

We'll assume all states are equally likely (cf earlier discn of microcanonical distrib), so we're using a fair coin.

Let $\Omega(E)$ = # of ways of getting $H = E$. (Up to normalization, this is the "measure" of the assoc set in phase space, using counting measure.) By CLT, $s_1 + \cdots + s_N$ is approx Gaussian as $N \to \infty$, with mean 0 + variance N, so

$$\Omega(E) \simeq C_N \, e^{-E^2/2N}$$

when E is not too large (at most $\sim \sqrt{N}$) and N is large. Evidently: due to redundancy (many ways to get same result) $H = E$ is vastly more likely if E is near 0 than if E is far from 0.

It is convenient to discuss

$$S(E) = \ln \Omega(E)$$

(abusing notation — this is different from the "informational entropy" discussed earlier). This is the "<u>microcanonical entropy</u>." Note that if we break system into 2 indep parts, # states is multiplicative but "entropy is additive", as I now explain.

$$s_1 \circ \circ \circ \circ \circ \circ \circ \, s_N$$

$$\underbrace{\qquad\qquad}_{\text{group } A} \quad \underbrace{\qquad\qquad}_{\text{group } B}$$

state X_A — state X_B

Hamiltonian H_A — Hamiltonian H_B

$$H(X) = H_A(X_A) + H_B(X_B) = \sum_{j=1}^{N} s_j$$

$$\text{when } X = (X_A, X_B).$$

<u>Question</u>: Given that $H(X) = E$, what is the most likely value of $E_A = H(X_A)$?

<u>Ans</u>: want $\max_{E_A} \Omega_A(E_A) \cdot \Omega_B(E - E_A)$

(by independence), or equivalently (taking log)

$$\max_{E_A} S_A(E_A) + S_B(E - E_A)$$

If N is large we can treat E_A as cont's var $\Rightarrow$ max when

$$S_A'(E_A^*) = S_B'(E - E_A^*)$$

Convention: $S_A'(E_A^*) = \beta = \frac{1}{T}$ (β = inverse temperature).

(Units: entropy is nondim'l here, so T has units of energy. If we measured temp in degrees we would need a conversion factor — Boltzmann's const k_B, giving $\beta = \frac{1}{k_B(\text{temp})}$.)

Recall our discn using central limit thm:

$$\Omega_A(E_A) = \text{const} \cdot e^{-E_A^2/2N_A}$$

$$\Rightarrow S_A(E_A) = \frac{-E_A^2}{2N_A} + \text{const}$$

$$\Rightarrow S_A'(E_A) = -E_A/N_A$$

so the most likely state has

$$\frac{-E_A}{N_A} = \frac{-E_B}{N_B} \qquad E_A + E_B = E.$$

<u>Another question</u>: given that $H(X) = E$, find

a) the <u>conditional probability</u> of event X_A?
b) the prob that $H_A(X_A) = E_A$?

Ans to (a): $$\text{Prob}(X_A \mid H = E) = \frac{\Omega_B(E - H_A(X_A))}{Z(E)}$$

$$\text{where } Z(E) = \sum_{\substack{\text{all states} \\ X_A \text{ of system } A}}' \Omega_B(E - H_A(X_A))$$

since if X_A is fixed then X_B must have $H_B(X_B) = E - H_A(X_A)$ and all states with this property are equally likely.

Ans to (b):

$$\text{Prob}\left(H(X_A) = E_A \mid H = E\right) = \frac{\Omega_A(E_A)\,\Omega_B(E - E_A)}{Z(E)}.$$

<u>Now here's the main pt.</u> Suppose we're interested in considering a <u>small</u> system (in mechanics: one particle) as part of a much larger system (eg many interacting particles). We'll model this by keeping system A fixed but taking size of B to ∞ (B is "the reservoir").

Let's revisit our formula for cond prob with this in mind:

$$\text{Prob}\{X_A \mid H = E\} = \frac{1}{Z(E)}\, e^{S_B(E - E_A)}, \quad \text{if } E_A = H_A(X_A)$$

Use lin approxn of entropy

$$S_B(E - E_A) \simeq S_B(E) - S_B'(E)\, E_A$$

so

$$\text{Prob}\{X_A \mid H = E\} \simeq c_E\, e^{-\beta E_A}$$

$$\text{with } \beta = S_B'(E)$$

Evidently: holding E fixed, but focusing on system A and dropping the cond expectation notation,

$$\text{Prob of state } X_A = \frac{1}{Z(\beta)}\, e^{-\beta H(X_A)}$$

$$\text{where } Z(\beta) = \sum_{\substack{\text{states } X_A \\ \text{of system } A}} e^{-\beta H(X_A)}$$

<u>This is the canonical distrn!</u>

Simple arithmetic reveals that under this distrn,

$$\langle H \rangle = \frac{1}{Z} \sum_X H(X)\, e^{-\beta H(X)} = -\frac{\partial}{\partial \beta} \ln Z$$

and $$\text{var}(H) = \langle H^2 \rangle - \langle H \rangle^2 = + \frac{\partial^2}{\partial \beta^2} \ln Z$$