Lecture 1:

Discuss course outline and pass out sign-up sheet

NOT a full course in any one of the four subjects.

Course focuses on ideas in statistical mechanics imported to ergodic theory.

Pre-reqs: functional analysis: L^p spaces, duality between measures and continuous functions, some fixed point theorems.

Evaluation: Two problem sets (not to be turned in but to be presented by students in review sessions) and a talk on topic of your choice.

Don't worry about diff, geom. for topic 4.

Notes will be posted.

Course time?

ENTROPY:

Defn: The *entropy* of a probability vector $\overline{p} = (p_1, \ldots, p_n)$ is

$$H(\overline{p}) := -\sum_{i=1}^{n} p_i \log p_i$$

Base of log is irrelevant, just changes by a multiplicative constant; ergodic theory assumes base is e (information theory assumes base is 2).

We take: $0 \log 0 = 0$, since $\lim_{x \to 0} x \log x = 0$

Prop:

0) $H(\overline{p}) \geq 0$
1) $H(\overline{p}) = 0$ iff $\overline{p}$ is deterministic, i.e., for some unique i, $p_i = 1$.
2) $H(\overline{p})$ is continuous and strictly concave

3) $H(\overline{p}) \leq \log n$ with equality iff $\overline{p}$ is uniform, i.e., $\overline{p} = (1/n, \ldots, 1/n)$

Proof: $H(\overline{p}) = \sum_{i=1}^{n} f(p_i)$ where $f(x) := -x \log x$,
0): $f(x) \geq 0$ on $[0, 1]$.

1): $f(x) = 0$ iff $x = 0$ or 1.

2): $f(x)$ is continuous and strictly concave since $f''(x) = -1/x < 0$.

3): Apply Jensen to f:

$$(1/n)H(\overline{p}) = \sum_{i=1}^{n}(1/n)f(p_i) \leq f(\sum_{i=1}^{n}(1/n)p_i) = f(1/n) = \frac{\log n}{n}$$

So,

$$H(\overline{p}) \leq \log(n)$$

with equality if $p_i = 1/n$ (and only if by strict concavity). $\square$

Another viewpoint:

Entropy of a random variable X with *finite* range $\{x_1, \ldots, x_n\}$: $X \sim p(x)$: i.e., $p_i = p(X = x_i)$:

$$H(X) := H(\overline{p})$$

where $\overline{p} = (p_1, \ldots, p_n)$

Note: $H(X)$ does not depend on values of X, but only on its distribution $\overline{p}$.

Intuitive meaning: $H(X)$ represents

— uncertainty in outcomes of X
— information gained in revealing an outcome of X.
— degree of randomness or disorder of X

Lecture 2:

Recall: For a probability vector $\overline{p} = (p_1, \ldots, p_n)$,

$$H(\overline{p}) := -\sum_{i=1}^{n} p_i \log p_i$$

Subject to some simple natural axioms, there is really only one choice for H (up to the base of the logarithm).

Can also define entropy for countable state probability vectors and, more generally continuous-valued probability distributions, but not in this course.

STATISTICAL MECHANICS: Boltzmann, Gibbs,

Ideal gas

Micro-state: physical state of system at a given time, e.g., positions and momenta of all particles or clustering of such vectors

Macro-state: probability distribution on set of micro-states

Laws of Thermodynamics:

– 1st law of thermo: without any external influences, energy of macro-state is fixed

– 2nd law of thermo: the macro-state tends to a state of maximal disorder, i.e., maximum entropy, subjected to fixed energy; such a state is called an *"equilibrium state"*. In an equilibrium state there are no macroscopic changes but the micro-states can evolve.

TRY to make this precise:

Let $\{s_1, \ldots, s_n\}$ be the collection of micro-states.
Let $u_i = U(s_i)$ be the energy of s_i (here, U is the energy function)
Let E^* be a fixed value of energy of the system.

FIND

$$\arg \max_{E_{\overline{p}}U=E^*} H(\overline{p})$$

Constrained optimization problem:

Maximize $H(\overline{p})$ subject to

$\sum_i u_i p_i = E^*$
$\sum_i p_i = 1$
$p_i \geq 0.$

Apply Lagrange multipliers:
grad $H = \beta U + \lambda \mathbf{1}$ (viewed as vector equation)

$$-\log p_i - 1 = \beta u_i + \lambda$$

(assuming base $= e$)

Solution:

$$p_i = ce^{-\beta u_i} = \frac{e^{-\beta u_i}}{Z(\beta)}$$

where Z is the normalization factor:

$$Z(\beta) = \sum_j e^{-\beta u_j}.$$

Call this prob dist: μ_β. $\square$

A probability distribution of the form μ_β above is called a *Gibbs state* – defined only by thermodynamics quantities U and β. So, *every equilibrium state is an (explicit) Gibbs state.*

Can show that there is a unique $\beta = \beta^* \in (-\infty, \infty)$ such that

$$E_{\mu_{\beta^*}}(U) = E^*$$

and this choice uniquely achieves global max (assuming that $\min U < E^* < \max U$):

Onto:
$\lim_{\beta\to+\infty} E_{\mu_\beta}(U) = \min U$
$\lim_{\beta\to-\infty} E_{\mu_\beta}(U) = \max U$

1-1: Compute $\frac{dE_{\mu_\beta}(U)}{d\beta} = -Var(\mu_\beta) < 0$
Observe:

$$H(\mu_{\beta^*}) = -\sum_i p_i \log \frac{e^{-\beta^* u_i}}{Z(\beta)} = \log Z(\beta^*) + \beta^* E_{\mu_{\beta^*}}(U).$$

Re-write:

$$(-1/\beta^*) \log Z(\beta^*) = E_{\mu_{\beta^*}}(U) - (1/\beta^*) H(\mu_{\beta^*})$$

In stat. mech., interpreted as "free energy" = "internal energy - (temperature)(entropy)"

A normalization factor $Z(\beta^*)$ and a Lagrange multiplier β^* have physical meaning.

A goal of this course: to show that these ideas hold in great generality, with general defns of equilibrium state and Gibbs state.

ERGODIC THEORY:

Defn: A measure-preserving transformation (MPT) on a *probability* space $(M, \mathcal{A}, \mu)$ is map $T : M \to M$ which is measurable and measure-preserving, i.e.,

For $A \in \mathcal{A}$, $T^{-1}(A) \in \mathcal{A}$ and $\mu(T^{-1}(A)) = \mu(A)$.

Example 1: Doubling map (w.r.t. normalized Lebesgue measure on the unit inteval)

$M = [0, 1)$, $\mathcal{A} =$ Borel σ-algebra, $\mu =$ Lebesgue measure
$T(x) = 2x \bmod 1$
Draw graph, which has two pieces of slope 2:

Inverse image of an interval I is the union of two intervals each with length $(1/2)\ell(I)$.

Note: T would not be an MPT if we required $\mu(A) = \mu(TA)$.

(on the circle, the map is $z \mapsto z^2$)

Defn: An *Invertible MPT* (IMPT) is an MPT T which is a bijection a.e. and T^{-1} is an MPT.

Example 2: Circle rotation (w.r.t. normalized Lebesgue measure on the circle)

$M : [0, 2\pi)$ with normalized Lebesgue measure

$T_\alpha(\theta) = \theta + \alpha \bmod 2\pi$, $\alpha \in M$.

Graph has slope 1.

MPT because Lebesgue measure is translation invariant.

Can also be viewed as map from circle to itself (the map is $z \mapsto e^{i\alpha} z$)

More examples soon.

Why measure is required to be preserved under *inverse* images.

Proposition: T is an MPT iff $\int f \circ T d\mu = \int f d\mu$ for all $f \in L^1$.

Proof:

If: Apply to $f = \chi_A$, $A \in \mathcal{A}$ and note that $\chi_A \circ T = \chi_{T^{-1}(A)}$:

$$\mu(T^{-1}(A)) = \int \chi_{T^{-1}(A)} d\mu = \int \chi_A \circ T d\mu = \int \chi_A d\mu = \mu(A).$$

Only If: We have $\int \phi \circ T d\mu = \int \phi d\mu$ for all simple ϕ.

By splitting f into positive and negative parts, we may assume $f \geq 0$.

There exists a monotone increasing sequence of simple functions $\phi_n \uparrow f$ and thus $\phi_n \circ T \uparrow f \circ T$.

$$\int f \circ T d\mu = \lim \int \phi_n \circ T d\mu = \lim \int \phi_n d\mu = \int f d\mu. \square$$

Lecture 3:

Will post lectures at the end of each week.

Will post a running list of exercises (for review session in early February).

Recall Proposition: T is an MPT iff $\int f \circ T d\mu = \int f d\mu$ for all $f \in L^1$.

Remark: same result holds for L^p for all $1 \leq p \leq \infty$ since on a probability space $L^p \subset L^1$:

$\int |f| d\mu = \int_{|f| \leq 1} |f| d\mu + \int_{|f|>1} |f| d\mu \leq 1 + \int |f|^p d\mu$

Allows to define a unitary operator $U_p : L^p \to L^p$ by $f \mapsto f \circ T$ (of special interest is $p = 2$).

Can verify MP on a small, concrete sub-collection of $\mathcal{A}$:

Defn: A *semi-algebra* $\mathcal{B}$ on a set M is a collection of sets s.t.

1. $\emptyset, M \in \mathcal{B}$
2. $\mathcal{B}$ is closed under finite intersections
3. for $B \in \mathcal{B}$, B^c is a finite *disjoint* union of elements of $\mathcal{B}$.

Note: this is a weaker concept than an algebra

Examples:

- Intervals in $\mathbb{R}$,
- (literal) rectangles in $\mathbb{R}^2$
- cylinder sets $F^{Z^+} := \{x_0 x_1 x_2 \cdots : \quad x_i \in F\}$ (F is a finite alphabet):

$$A = \{x \in F^{Z^+} : \ x_{i_1} = a_1 \ldots x_{i_k} = a_k\}$$

Defn: A semi-algebra $\mathcal{B}$ *generates* a σ-algebra $\mathcal{A}$ if $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{B}$.

Examples: Borel sets in $\mathbb{R}$, $\mathbb{R}^2$ and Borel sets in the product σ-algebras F^{Z^+} or $F^Z := \{\ldots x_{-2}x_{-1}x_0x_1x_2\cdots : \quad x_i \in F\}$.

Theorem: Let $\mathcal{B}$ be a generating semi-algebra for $\mathcal{A}$. Then T is MPT iff for all $B \in \mathcal{B}$, $T^{-1}(B) \in \mathcal{A}$ and $\mu(T^{-1}(B)) = \mu(B)$

Proof:

Only if: obvious

If: argue using monotone class lemma:

Let

$$\mathcal{C} = \{A \in \mathcal{A} : T^{-1}(A) \in \mathcal{A} \text{ and } \mu(T^{-1}(A)) = \mu(A)\}$$

Then $\mathcal{C}$ contains $\mathcal{B}$ and hence the algebra generated by $\mathcal{B}$ (the algebra is the set of all finite *disjoint* unions of elts. of $\mathcal{B}$; it is a algebra because it is closed under finite intersections and complements)

And $\mathcal{C}$ is a monotone class, (i.e., closed under countable increasing sequence of s sets and countable decreasing sequences of sets).

Then $\mathcal{C}$ contains $\mathcal{A}$, by Monotone class lemma. $\square$

Examples: Check MPT on semi-algebra:

1. Doubling map: check on intervals
2. Rotation on the circle: check on intervals
3. Baker's transformation

M : unit square with Lebesgue measure

$T(x, y) = (2x \bmod 1, (1/2)y) \quad \text{if } 0 \leq x < 1/2$

$T(x, y) = (2x \bmod 1, (1/2)y + 1/2) \quad \text{if } 1/2 \leq x < 1$

Draw inverse image of a rectangle contained in bottom or top. Inverse image has half the width and twice the height.

If a rectangle intersects top and bottom, then split it into top part and bottom part.

4. One-sided stationary process (finite range) e.g., iid or stationary Markov $M = F^{\mathbb{Z}^+}$ (F is a finite alphabet)

$\mathcal{A}$: product σ-algebra using discrete σ-algebra on F.

for cylinder set

$$A = \{x \in M : x_{i_1} = a_1 \dots x_{i_k} = a_k\}$$

define

$$\mu(A) := p(X_{i_1} = a_1 \dots X_{i_k} = a_k)$$

where p is law of process. Extend μ to product sigma-algebra.

$T := \sigma$, the left shift map: $\sigma(x)_i = x_{i+1}$.

$$\sigma^{-1}(A) = \{x \in M : x_{i_1+1} = a_1 \dots x_{i_k+1} = a_k\}$$

$\mu(\sigma^{-1}(A)) = \mu(A)$, by stationarity (in fact, stationarity of the process is equivalent to measure-preserving of σ)

Note that sometimes $\mu(\sigma(A)) \neq \mu(A)$, e.g.,

$$A = \{x : x_0 = 1\}, \quad \sigma(A) = M, \text{ entire space.}$$

falls off cliff

5. Two-sided stationary process (finite range): processes on $\mathbb{Z}$, instead of $\mathbb{Z}^+$.

$M = F^{\mathbb{Z}}$

$T := \sigma$, the left shift map: $\sigma(x)_i = x_{i+1}$.

Later will study processes on Z^d and Z^d_+.

Poincare Recurrence Theorem (1890) ("Every MPT is recurrent"):

Let T be an MPT and $\mu(A) > 0$. Then a.e. $x \in A$ returns to A infinitely often, i.e, for a.e. $x \in A$, there are infinitely many $n > 0$ s.t. $T^n(x) \in A$.

Proof:

Lemma: There exists $n > 0$ such that $\mu(T^{-n}(A) \cap A) > 0$.

Proof: If not, we claim that $\{T^{-n}(A)\}_{n=0}^{\infty}$ are pairwise measure disjoint (i.e., for $n \neq m$, $\mu(T^{-n}(A) \cap T^{-m}(A)) = 0$); this follows because (assuming $n > m$):

$$T^{-n}(A) \cap T^{-m}(A) = T^{-m}(T^{-(n-m)}(A) \cap A)$$

So,

$$\mu(T^{-n}(A) \cap T^{-m}(A)) = 0.$$

Thus, $\mu(\cup_{n=0}^{\infty} T^{-n}(A)) = \infty$. $\square$

Let G be the set of points in A which do not visit A infinitely often. We want to show that $\mu(G) = 0$.

Let G_n be the set of points *in* A that visit A exactly n times in the future.

Then G is the disjoint union of the G_n and so it suffices to show that $\mu(G_n) = 0$.

If $\mu(G_n) > 0$, then by Lemma, there exists $n' > 0$ s.t. $\mu((T^{-n'}(G_n) \cap G_n)) > 0$.

If x is in this intersection, then $x \in G_n$ and $T^{n'}(x) \in G_n$ and so x visits A at least $n + 1$ times, a contradiction to defn. of G_n. $\square$