2. Entropy

2. Measure-theoretic entropy

For $(X, \mathcal{B}, \mu)$ probability space, the entropy of a partition ξ is given by

$$H(\xi) = H_\mu(\xi) = -\sum_{C \in \xi} \mu(C) \log \mu(C)$$

$0 \log 0 := 0$

Remarks: 1) $H(\xi) \geq 0$

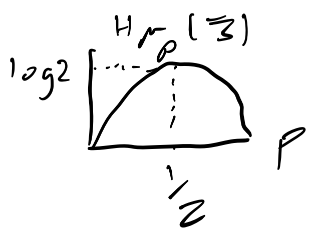
2) $H(\xi) = 0 \iff \exists C \in \xi$ with $\mu(C) = 1$

3) If $\xi = \{A_1, \dots, A_k\}$, and $\mu(A_i) = \frac{1}{k}$ for each i, then $H(\xi) = -\sum_{i=1}^{k} \frac{1}{k} \log \frac{1}{k} = \log k$

largest possible entropy for partition into k sets

$\xi = \{A_1, A_2\}$

$\mu_p(A_1) = p$



Defn The information function is

$$I(\xi)(x) := -\log(\mu(\xi(x)))$$

$$= \sum_{A \in \xi} -\log(\mu(A))\, I_A(x)$$

Note ($I(\xi)$ is a measurable fn.) and $I(\xi) \geq 0$

Note that we can write

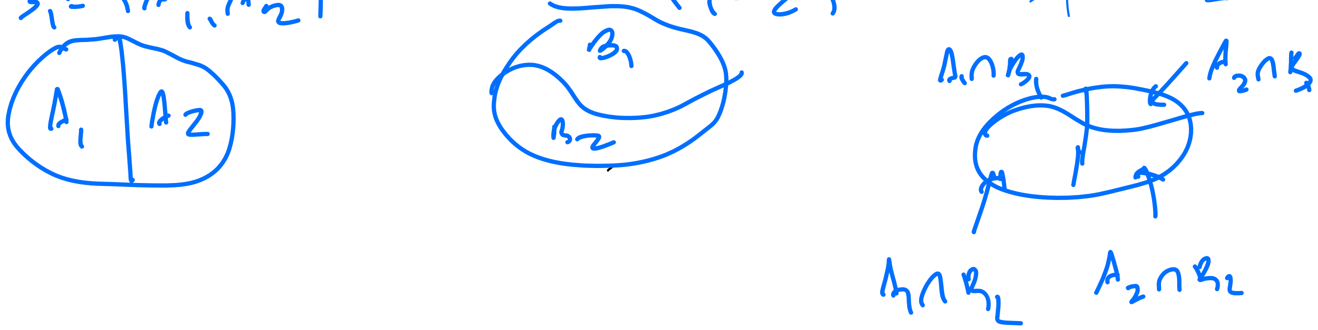
$$H(\xi) = \int I(\xi)\, d\mu$$

i.e. $H(\xi)$ is $\boxed{\text{average information}}$

For finite partitions $\xi_0, \dots, \xi_{n-1}$, the partition $\bigvee\limits_{i=0}^{n-1} \xi_i$ is determined by:

$$\bigvee_{i=0}^{n-1} \xi_i(x) = \bigcap_{i=0}^{n-1} \xi_i(x).$$

$\xi_1 = \{A_1, A_2\}$ $\quad\quad$ $\xi_2 = \{B_1, B_2\}$ $\quad\quad$ $\xi_1 \vee \xi_2$



coarser $\quad\quad$ finer

We write $\xi \leq \eta$ if each element of η is contained in an element of ξ (mod 0)

Good notation because $\#\xi \leq \#\eta$

Interpretation

"Guess who?" game. Ask questions to find opponent's identity.

Each Yes/No question gives a partition

Denote $I(p)$ as the information received if p is the prob. of a positive answer

- Claim: $I(p) = -\log p$ is the natural function that meets criteria to be information:

Want:
- $I(p) > 0$ when $p > 0$
- Choosing a smaller set gives more information
- Information gained by independent results sums

But prob. of positive answer p twice is p^2. So want $I(p^2) = 2I(p)$

Qn 1: "Is your character female" is 50-50 in the game

Qn 2: "Is your character bald?" gives a lot of info if yes, not much if no.

Entropy is the average "information"

Qn 1 gives more info on average.

Suppose $X = \{x_1, \dots, x_n\}$ with respective probabilities $(p_1, \dots, p_n)$, $\sum p_i = 1$

p_i is the prob you pick character i to be your card

[e.g. Guess who. A good player will choose in a uniform way but maybe a player will favor choosing women or characters with glasses)

Think of a question as a partition

A sequence of questions $\xi_1, \dots, \xi_n$ is generating if they can distinguish all results

partition into points

i.e. $\hat{\bigvee_{i=1}} \xi_i(x) = \{x\}$

By definition, $H\left(\bigvee_{i=1}^{\infty} \xi_i\right) = -\sum p_i \log p_i$

Let $h(\mu) = H\left(\bigvee_{i=1}^{\infty} \xi_i\right)$

for any generating sequence.
A priori, the average information could depend on the (ξ_i). It doesn't, by simple computation. So now we can talk about entropy of a measure, not a partition.

measure entropy
For a measure preserving transformation $T: X \to X$

$$T^{-1}\xi := \{T^{-1}A : A \in \xi\}$$

Defn If ξ is a finite partition of $(X, \mathcal{B}, \mu)$ then the entropy of T w.r.t. ξ is

$$h(\mu, T, \xi) := \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right)$$

Limit exists because $a_n = H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$ is sub-additive. Limit equals $\inf_n \frac{a_n}{n} \leq a_1 = H(\xi) < \infty$

<u>Defn</u> The entropy of $(X, \mathcal{B}, \mu, T)$ is

$$h(\mu) = \sup_{\xi} h(\mu, T, \xi)$$

where sup is over all <u>finite</u> partitions

Want to find ξ so $h(\mu) = h(\mu, T, \xi)$

it can be checked that if $\xi_n = \bigvee_{i=0}^{n-1} T^{-i}\xi$ or $\bigvee_{-n}^{n} T^{-i}\xi$

satisfies $\operatorname{diam}(\xi_n) \to 0$ i.e. $\bigcap_{n=1}^{\infty} \xi_n(x) = \{x\}$ $\forall x$

i.e. if the sequence of "questions" ξ_n (given by improving ξ using the dynamics) sees every outcome, the ξ gives the entropy

For a flow $F = (f_t)$ it turns out

$$h(\mu, f_t) = |t|\, h(\mu, f_1) \quad \text{(Abramov formula)}$$

So we let $h(\mu) := h(\mu, f_1)$

2.2 Partitions and symbolic dynamics

$X \subseteq \{0, 1, \dots, k\}^{\mathbb{Z}}$ closed $\quad \sigma(X) = X$

$\sigma : X \to \{0, \dots, k\}^{\mathbb{Z}}$ shift map

$$x = \dots\dots\; x_{-3}\, x_{-2}\, x_{-1} . x_0\, x_1\, x_2\, x_3 \dots\dots$$

$$\downarrow \sigma$$

$$y = \sigma(x) = \dots\; x_{-2}\, x_{-1}\, x_0 . x_1\, x_2\, x_3\, x_4 \dots$$

For a shift space X,

$L = L(X)$: language of X

= the set of all finite words that occur in points of X

L_n = words in language of length n

Natural metric d on X:

$d(x,y) = \frac{1}{2^{|k|}}$ where k is first index x and y disagree

If $x_0 \neq y_0$, $d(x,y) = 1$
If $x_1 \neq y_1$, $d(x,y) = \frac{1}{2}$
etc..

cylinder sets: If $w = w_1 w_2 \dots w_k$

$${}_p[w] = \{y \in X : y_p \cdots y_{p+k-1} = w_1 \cdots w_k\}$$

so e.g. ${}_0[i] = \{y \in X : 0^{th}$ entry is $i.\}$

Partitions

Let $\xi = \xi_0 = \{ {}_0[0], {}_0[1], \ldots, {}_0[k] \}$

Let $\xi_n = \bigvee_{i=-n}^{n} \sigma^{-i} \xi_0$

Sets $A \in \xi_n$ are cylinder sets of form

$${}_{-n}[x_{-n} x_{-n+1} \ldots x_n]$$

$\mathrm{Diam}(\xi_n) \leq \frac{1}{2^{n+1}} \to 0$

Thus for $\underline{\text{any}}$ σ-inv. prob measure

Thus $h(\mu, \sigma) = h(\mu, \sigma, \xi)$

$$= \lim_{n \to \infty} \frac{1}{n} H_\mu \left(\bigvee_{i=0}^{n-1} \sigma^{-i} \xi \right)$$

$$\boxed{h(\mu, \sigma) = \lim_{n \to \infty} -\frac{1}{n} \sum_{w \in L_n} \mu(w) \log \mu(w)}$$

where $\mu(w) := \mu({}_0[w]) = \mu({}_p[w]) \;\; \forall p \in \mathbb{Z}$

since μ is invariant

Topological entropy of shift space is

$$h_{top}(\sigma) = \lim_{n \to \infty} \frac{1}{n} \log \# L_n$$

2.3 Topological entropy, etc

General case (X, f). X compact metric spa[ce]
$f: X \to X$ cb

Defn A <u>Bowen ball</u> is a set of the form

$$B_n(x, \varepsilon) = \{ y \in X : d(f^i x, f^i y) < \varepsilon \ \forall i \in \{0, \dots, n-1\} \}$$

Note: for (Σ, σ), the Bowen balls are cylinder sets

(Let $d_n(x, y) = \max\{ d(f^i x, f^i y) : i \in \{0, \dots, n-1\} \}$
Then d_n is a metric. Bowen balls are just d_n-metric balls)

E is an <u>(n, ε)-spanning set</u> if

$$\bigcup_{x \in E} \overline{B}_n(x, \varepsilon) = X$$

Let $r_n(\varepsilon, X) = \min\{ \# E : E \text{ is an } (n, \varepsilon)\text{-spanning set} \}$

Let $\boxed{h_{top}(f, \varepsilon) := \overline{\lim} \frac{1}{n} \log r_n(\varepsilon, X)}$

Let $\boxed{h_{top}(f) = \lim_{\varepsilon \to 0} h_{top}(f, \varepsilon)}$

Katok entropy formula for $f: X \to X$

Let $Z \subset X$.

Let $r_n(\varepsilon, Z) = \min\{\# E : E \text{ is } (n,\varepsilon)\text{-spanning for } Z\}$

i.e. $Z \subseteq \bigcup_{x \in E} \overline{B_n(x,\varepsilon)}$

ergodic inv. prob.

Let $\mu \in M^e_f(X)$. For a fixed $\delta \in (0,1)$ (e.g. $\delta = \frac{1}{2}$), then

Thm

$$h(\mu, f) = \lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}} \frac{1}{n} \log\left(\inf\{r_n(\varepsilon, Z) : Z \text{ has } \mu(Z) > 1-\delta\}\right)$$

sketch:

Comes from local entropy formula:

for μ-a.e. x $\quad \mu(B_n(x,\varepsilon)) \approx e^{-nh(\mu)}$

Brin-Katok / Shannon-McMillan-Breiman in symbolic setting

$\therefore$ Need $\approx e^{nh(\mu)}$ such sets to cover a set of measure $1-\delta$.

Note: Clear from Katok entropy formula that

$$(*) \qquad \sup_{\mu \in M^e_f(X)} h(\mu, f) \leq h_{top}(f)$$

Recall: Our goal is to study μ which achieve equality in $(*)$