Lecture 9.5

Entropy

1. Entropy of partitions

Let X be a set and let $\mathcal{B}$ = sigma algebra.

Defn. A partition $\alpha = \{A_1, \dots, A_k\}$ is a finite collection of m'ble sets such that:

1. $X = A_1 \cup \dots \cup A_k$;
2. $A_i \cap A_j \neq \emptyset$ for $i \neq j$.

Defn. We define the entropy of a finite partition α by:

$$H_\mu(\alpha) = -\sum_{i=1}^{k} \mu(A_i) \log \mu(A_i)$$

NB. This quantifies the information given by knowing which elt of α a point lies in

Defn. Let $\left\{\begin{array}{l}\beta = \{B_1, \dots, B_\ell\} \\ \alpha = \{A_1, \dots, A_k\}\end{array}\right\}$ be two partitions of X. We define the conditional entropy by:

$$H_\mu(\beta|\alpha) = -\sum_{i=1}^{k}\sum_{j=1}^{\ell} \mu(A_i \cap B_j) \log \left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)$$

Remark: If $\alpha = \{X\}$ then $H_\mu(\beta|\alpha) = H_\mu(\beta)$

Defn. Assume that $\alpha = \{A_i\}$, $\beta = \{B_j\}$ are partitions and we define the refinement:

$$\alpha \vee \beta = \{A_i \cap B_j\}$$

<u>Proposition.</u>

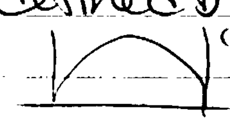
1. $H_\mu(\alpha \vee \beta) = H_\mu(\alpha) + H_\mu(\beta|\alpha)$;
2. $H_\mu(\beta|\alpha) \leq H_\mu(\beta)$;
3. $H_\mu(\alpha) = H_\mu(T^{-1}\alpha)$, where $T^{-1}\alpha = \{T^{-1}A_1, \dots, T^{-1}A_n\}$

<u>Proof</u> 1. $H_\mu(\alpha \vee \beta) = -\sum_{i,j} \mu(A_i \cap B_j)\log \mu(A_i \cap B_j)$

$$= -\sum_{i,j} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)$$

$$- \sum_{i,j} \mu(A_i \cap B_j)\log \mu(A_i)$$

$$= H_\mu(\beta|\alpha) + H_\mu(\alpha)$$

2. Since the function $f : (0,1) \to \mathbb{R}$ defined by $f(t) = -t\log t$ is concave 

we can write:

$$H_\mu(\alpha|\beta) = -\sum_{i=1}^{k}\sum_{j=1}^{l} \mu(A_i \cap B_j)\log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)$$

$$= -\sum_{j=1}^{l}\left(\sum_{i=1}^{k} \mu(A_i)\underbrace{\left[\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)\right]}_{f\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)}\right)$$

$$\leq \sum_{j=1}^{l} f\left(\sum_{i=1}^{k} \cancel{\mu(A_i)}\left[\frac{\mu(A_i \cap B_j)}{\cancel{\mu(A_i)}}\right]\right) = H(\beta)$$

3. Since $\mu(T^{-1}A_i) = \mu(A_i)$ we see that

$$H_\mu(T^{-1}\alpha) = -\sum_{i=1}^{k} \mu(T^{-1}A_i)\log \mu(T^{-1}A_i)$$

$$= -\sum_{i=1}^{k} \mu(A_i)\log \mu(A_i) = H_\mu(\alpha).$$

<u>Cor</u> $H_\mu(\alpha \vee \beta) \leq H_\mu(\alpha) + H_\mu(\beta)$.

<u>Proof</u> We have by the above:

$$H_\mu(\alpha \vee \beta) \underset{\uparrow \text{part 1}}{=} H_\mu(\alpha) + \underbrace{H_\mu(\beta|\alpha)}_{\underset{\uparrow \text{part 2.}}{\leq H_\mu(\beta)}} \leq H_\mu(\alpha) + H_\mu(\beta) \quad \square$$

By induction we can define n-fold refinement and:

$$\alpha^{(n)} := \bigvee_{i=0}^{n-1} T^{-i}\alpha$$

We then have:

<u>Lemma</u>: For $n, m \geq 1$ we have:

$$H_\mu(\alpha^{(n+m)}) \leq H_\mu(\alpha^{(n)}) + H_\mu(\alpha^{(m)})$$

<u>Proof</u> By the corollary:

$$H_\mu(\alpha^{(n+m)}) = H_\mu(\alpha^{(n)} \vee T^{-n}\alpha^{(m)})$$

$$\underset{\text{By Cor}}{\leq} H_\mu(\alpha^{(n)}) + \underbrace{H_\mu(T^{-n}\alpha^{(m)})}_{= H_\mu(\alpha^{(m)})}$$

$$\leq H_\mu(\alpha^{(n)}) + H_\mu(\alpha^{(m)}).$$

<u>Def^n</u>. A sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ are called subadditive if $a_{n+m} \leq a_n + a_{\cancel{n}m}$ for $n, m \geq 1$. (Otherwise, it's definition of nonnegative sequence)

The main application is the following:

Lemma (Subadditive sequence lemma).
If $(a_n)_{n=1}^{\infty}$ is a subadditive sequence then

$$\lim_{n\to\infty} \frac{a_n}{n} = l := \inf_{n\geq 1}\left\{\frac{a_n}{n}\right\}$$

Proof. Given $\varepsilon > 0$, we can choose $N > 0$ such that $\frac{a_N}{N} < l + \varepsilon$.

For any $n \geq 1$ we can write $n = lN + r$ $\begin{cases} l \geq 0 \\ 0 \leq r \leq N-1 \end{cases}$

By subadditivity:

$$a_n \leq l a_N + a_r$$

Thus we can write:

$$\frac{a_n}{n} \leq \frac{l a_N}{n} + \frac{a_r}{n} \leq \underbrace{\frac{a_N}{N + r/l}}_{\xrightarrow[l\to\infty]{} \frac{a_N}{N}} + \underbrace{\frac{\max\limits_{0\leq r\leq N-1} a_r}{n}}_{\xrightarrow[n\to\infty]{} 0}$$

$$\leq l + 2\varepsilon \text{ for } n, l \text{ sufficiently large.}$$

Since $\varepsilon > 0$ is arbitrary, the result follows $\square$

We can apply the lemma to $a_n = H_\mu(\alpha^{(n)})$.

Definition: We can define the entropy of the partition α w.r.t T and μ:

$$h_\mu(T, \alpha) = \lim_{n\to\infty} \frac{H_\mu(\alpha^{(n)})}{n}$$

Example 1 (Doubling map).
Let $T: \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by $Tx = 2x \pmod 1$.
Let $\alpha = \{[0, 1/2), [1/2, 1)\}$
Let μ = Lebesgue measure.
For $n \geq 1$, $\alpha^{(n)} = \bigvee_{i=0}^{n-1} T^{-i}\alpha$ consists of intervals $\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)$ for $i = 0, \ldots, 2^{n+1}$ (each with Lebesgue meas. $1/2^{n+1}$)

Thus we have that

$$H_\mu(\alpha^{(n)}) = -2^{n+1} \cdot \left(\frac{1}{2^{n+1}}\right) \log\left(1/2^{n+1}\right) = (n+1)\log 2$$

and therefore $h_\mu(T) = \lim_{n\to\infty} \frac{H_\mu(\alpha^{(n)})}{n} = \log 2$.

Entropy for transformations
We can define the entropy for $T: X \to X$ and a T-invariant ~~measure~~ μ.

Defn We define the entropy of $T: X \to X$ (with respect to μ) by:

$$h_\mu(T) = \sup\{h_\mu(T, \alpha) \mid \alpha = \text{finite partition}\}$$

Remark. This value may turn out to be infinite.

Let $T_1: X_1 \to X_1$ preserve a probability measure μ_1 and let $T_2: X_2 \to X_2$ preserve a probability measure μ_2.

Defn. We define an isomorphism $\pi : X_1 \to X_2$ to be a bijection a.e., such that:

(a) π, π^{-1} are m'ble;

(b) $T_2 \circ \pi = \pi \circ T_1$;

(c) $\pi_* \mu_1 = \mu_2$

The main result on entropy is the following.

Theorem (Kolmogorov-Sinai). The entropy is an isomorphism invariant.

Proof. Let $\alpha_1 = \{A_1, \ldots, A_n\}$ is a partition for X_1 then $\pi(\alpha_1) = \{\pi(A_1), \ldots, \pi(A_n)\}$ is a partition for X_2.

From the definitions: $h_{\mu_1}(T_1, \alpha_1) = h_{\mu_2}(T_2, \pi(\alpha_1))$

In particular, $h_{\mu_1}(T_1) \leqslant h_{\mu_2}(T_2)$.

Conversely, if $\alpha_2 = \{A_1, \ldots, A_n\}$ is a partition for X_2 then $\pi^{-1}\alpha_2 = \{\pi^{-1}(A_1), \ldots, \pi^{-1}(A_n)\}$ is a partition for X_1.

From the definitions, $h_{\mu_2}(T_2, \alpha_2) = h_{\mu_1}(T_1, \pi^{-1}(\alpha_2))$.

In particular, we see that $h_{\mu_1}(T_1) \geqslant h_{\mu_2}(T_2)$.

Thus: $h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$, as required $\square$

Entropy from generators

Def^n. A generator is a finite partition β such that $\mathcal{B} = \bigvee_{k=0}^{\infty} T^{-k}\beta$. This means that $\forall A \in \mathcal{B}$, $\forall \varepsilon > 0$, $\exists N > 0$, $\exists B \in \left(\bigvee_{k=0}^{N-1} T^{-k}\beta\right)$ i.e., a finite union of elts of $\bigvee_{k=0}^{N-1} T^{-k}\beta$, such that

$$\mu(A \triangle B) < \varepsilon$$

(where $A \triangle B = (A-B) \cup (B-A)$)

Theorem. If β is a generator then $h_\mu(T) = h_\mu(T, \beta)$.

We need to show that if α is any finite partition then $h_\mu(T, \alpha) \leq h_\mu(T, \beta)$.

We begin with the following.

Def^n. Given two partitions α_1, α_2 we have $\alpha_1 < \alpha_2$ if the elts of α_1 are unions of elts of α_2. (In particular, $\#\alpha_1 \leq \#\alpha_2$).

Lemma (i) If $\alpha < \beta$ then $H_\mu(\alpha|\gamma) \leq H_\mu(\beta|\gamma)$

(ii) If $\alpha < \beta$ then $H_\mu(\gamma|\alpha) \geq H_\mu(\gamma|\beta)$

~~Proof. If we assume that $\alpha_1 < \alpha_2$ then $\alpha_2 = \alpha_1 \vee \alpha_2$.~~

~~Thus: $H_\mu(\alpha_2) = H_\mu(\alpha_1 \vee \alpha_2) = H_\mu(\alpha_1) + H_\mu(\alpha_2|\alpha_1)$~~

~~and the result follows.~~

(iii) $H_\mu(\alpha|\beta)$

$H_\mu(T^{-1}\alpha|T^{-1}\beta)$.

~~Proof of theorem:~~

~~1. By Lemma: $\alpha^{(n)} <$~~

Proof of lemma (i). $H_\mu(\beta|\gamma) = H_\mu(\alpha\vee\beta|\gamma)$
$$= H_\mu(\alpha|\gamma) + \underbrace{H_\mu(\beta|\gamma\vee\alpha)}_{\geq 0}$$
$$\geq H_\mu(\alpha|\gamma)$$

(ii). Recall $f(t) = -t\log t$ is concave.
For each $A_i \in \alpha$, $C_j \in \gamma$ observe that:
$$\sum_k \frac{\mu(A_i\cap B_k)}{\mu(A_i)} = 1.$$
Thus:
$$\left.\begin{aligned} & f\left(\sum_k \frac{\mu(A_i\cap B_k)}{\mu(A_i)}\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)\right) \\ & \geq \sum_k \frac{\mu(A_i\cap B_k)}{\mu(A_i)} f\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)\end{aligned}\right\} (*)$$
However, since $\alpha < \beta$ (each A_i union of sets from β):
$$\left.\sum_k \frac{\mu(A_i\cap B_k)}{\mu(A_i)} \cdot \frac{\mu(C_j\cap B_k)}{\mu(B_k)} = \frac{\mu(A_i\cap C_j)}{\mu(A_i)}\right\} (**).$$
Comparing $(*)$ & $(**)$:
$$-\frac{\mu(A_i\cap C_j)}{\mu(A_i)}\log\left(\frac{\mu(A_i\cap C_j)}{\mu(A_i)}\right) \geq -\sum_k \frac{\mu(A_i\cap B_k)}{\mu(A_i)}\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)\log\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)$$
Multiplying each side by $\mu(A_i)$ and summing gives:
$$\underbrace{-\sum_{ij}\mu(A_i\cap C_j)\log\left(\frac{\mu(A_i\cap C_j)}{\mu(A_i)}\right)}_{H_\mu(\gamma|\alpha)} \geq -\sum_{ij}\sum_k \mu(A_i\cap B_k)\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)\log\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)$$
$$= -\sum_j\sum_k \cancel{\mu(B_k)}\cdot\frac{\mu(C_j\cap B_k)}{\cancel{\mu(B_k)}}\log\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)$$
$$= \underbrace{-\sum_j\sum_k \mu(C_j\cap B_k)\log\left(\frac{\mu(C_j\cap B_k)}{\mu(B_k)}\right)}_{H_\mu(\gamma|\beta)}$$

Proof of theorem

1. Denote $\alpha^{(n)} = \bigvee_{i=0}^{n-1} T^{-i}\alpha$, $\beta^{(n)} = \bigvee_{i=0}^{n-1} T^{-i}\beta$

Then

$$H_\mu(\alpha^{(n)}|\beta^{(n)}) \leq \sum_{i=0}^{n-1} H_\mu(T^{-i}\alpha|\beta^{(n)}) \qquad (\text{by (i)})$$
$$\leq \sum_{i=0}^{n-1} H_\mu(T^{-i}\alpha|T^{-i}\beta) \qquad (\text{by (ii)})$$
$$= n H_\mu(\alpha|\beta) \qquad (\text{by (iii)})$$

Therefore we have that:

$$\frac{1}{n} H_\mu(\alpha^{(n)}) \leq \frac{1}{n} H_\mu(\beta^{(n)}) + \underbrace{\frac{1}{n} H_\mu(\alpha^{(n)}|\beta^{(n)})}_{< H_\mu(\alpha|\beta)}.$$

If we replace β by $\beta^{(N)} = \bigvee_{i=0}^{N-1} T^{-i}\beta$.

Then $\frac{1}{n} H_\mu(\alpha^{(n)}) \leq \frac{1}{n} H_\mu(\beta^{(n+N)}) + H(\alpha|\beta^{(N)})$

2. $H(\alpha|\beta^{(N)}) \to 0$ as $N \to +\infty$ [Exercise]

Thus: $h_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\alpha^{(n)})$

$$\leq \lim_{n\to\infty} \frac{1}{n} H_\mu(\beta^{(n+N)}) + H(\alpha|\beta^{(N)})$$
$$= \underbrace{h_\mu(T,\beta^{(N)})}_{h_\mu(T,\beta)} + \underbrace{H(\alpha|\beta^{(N)})}_{\to 0 \text{ as } N\to\infty \text{ [by 2.]}}$$

Example (Doubling map). For $\alpha = \{[0,\frac{1}{2}), [\frac{1}{2},1)\}$ we have $h_\mu(T,\alpha) = \log 2$. However, $\alpha^{(n)} = \bigvee_{i=0}^{n-1} T^{-i}\alpha$ consists of intervals $\{[\frac{i}{2^n}, \frac{i+1}{2^n}) \mid i=0,\dots,2^n-1\}$ which generate. Thus $h_\mu(T) = h_\mu(T,\alpha) = \log 2$

Proof of part (iii): $H_\mu(T^{-1}\alpha | T^{-1}\beta) = -\sum_{i,j} \mu(T^{-1}A_i \cap T^{-1}B_j) \log\left(\frac{\mu(T^{-1}A_i \cap T^{-1}B_j)}{\mu(T^{-1}B_j)}\right)$

$= -\sum_{i,j} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right)$

$= H_\mu(\alpha | \beta)$.

by T-invariance of μ.

<u>Example</u> (Rotations). Let $\rho \in \mathbb{R}$. Let $\begin{cases} T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z} \\ Tx = x + \rho \pmod 1 \end{cases}$

Let $\alpha = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$. Let $\mu =$ Lebesgue measure.

- If $\rho = p/q$ is irrational then $\alpha^{(n)} = \{[i/q, \frac{i+1}{q})\}$ and thus: $h_\mu(T) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\alpha^{(n)}) = 0.$

- If ρ is irrational then since $T^{-i}\alpha$ generates $\mathcal{B}$ we see that $H(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i}\alpha) \to 0$ as $n \to +\infty$.

Then $\frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) = \underbrace{\frac{1}{n} H_\mu(\alpha)}_{\to 0} + \underbrace{\frac{1}{n} H_\mu\left(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i}\alpha\right)}_{\to 0.}$

and we see:

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

In fact, the same is true for any finite partition

<u>Example</u> (Bernoulli shifts). Let $p = (p_1, \dots, p_k)$ and $\Sigma = \{1, \dots, k\}^{\mathbb{Z}}$ with shift $\sigma : \Sigma \to \Sigma$. ($0 < p_i < 1$ and $p_1 + \dots + p_k = 1$).

Let $\alpha = \{[1]_0, \dots, [k]_0\}$ be the partition by 1-cylinders. We then see that

$$\alpha^{(n)} = \bigvee_{i=0}^{n-1} \sigma^{-i}\alpha = \Big\{ \underbrace{[i_0, i_1, \dots, i_{n-1}]_0^{n-1}}_{n\text{-cylinder}} \,\Big|\, i_0, \dots, i_{n-1} \in \{1, \dots, k\} \Big\}$$

Then
$$\begin{aligned} H_\mu(\alpha^{(n)}) &= -\sum_{i_0 \cdots i_{n-1}} \mu([i_0, \dots, i_{n-1}]) \log \mu([i_0 \cdots i_{n-1}]) \\ &= -\sum_{i_0 \cdots i_{n-1}} \left(\prod_{j=0}^{n-1} p_{i_j}\right) \log\left(\prod_{j=0}^{n-1} p_{i_j}\right) \\ &= -n \sum_{j=1}^{k} p_j \log p_j \end{aligned}$$

Thus $h_\mu(\alpha) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) = -\sum_{i=1}^{k} p_i \log p_i$.

Example (Markov shift). Let $\sigma : \Sigma \to \Sigma$ be the subshift associated to a matrix A with entries 0 or 1. Then

$$\Sigma = \{ \underline{x} = (x_n) \in \{1, \dots, k\}^{\mathbb{Z}^+} \mid A(x_n, x_{n+1}) = 1, n \geq 0 \}$$

Let P be a stochastic matrix with $P(i,j) = 0$ iff $A(i,j) = 1$. Let $pP = p$ be left eigenvector.

Let $\alpha = \{ [i]_0 \mid i \in \{1, \dots, k\} \}$ be the partition of 1-cylinders.

Let

$$H_\mu(\alpha^{(n)}) = -\sum_{i_0 - i_{n-1}} \mu([i_0, \dots, i_{n-1}]) \log \mu([i_0, \dots, i_{n-1}])$$

$$= -\sum_{i_0 - i_{n-1}} p_{i_0} \prod_{j=0}^{n-1} P(i_j, i_{j+1}) \log\left(p_{i_0} \prod_{j=0}^{n-1} P(i_j, i_{j+1})\right)$$

$$= -\sum_{i=1}^{k} \sum_{j=1}^{k} p_i (\log P(i,j)) P(i,j)$$