2/11/2017.

Lecture 2:

reminder:

→ $(X, \mathcal{B}, \mu)$ is a Prob. space.

→ A partition $\xi = \{A_i \mid i \in I\}$, $X = \bigcup_{i \in I} A_i$

→ for two partitions $\xi = \{A_i \mid i \in I\}$, $\eta = \{B_j \mid j \in J\}$:

$\xi \vee \eta := \{A_i \cap B_j \mid \forall (i,j) \in I \times J\}$

→ the atom of x is $[x]_\xi = A_i$ where $x \in A_i$

→ $I_\mu(\xi) : X \longrightarrow [0,1]$

$$I_\mu(\xi)(x) = -\log(\mu([x]_\xi))$$

→ $H_\mu(\xi) = \int I_\mu(\xi)\, d\mu = -\sum_{i \in I} \mu(A_i) \log(\mu(A_i))$

Let $(X, \mathcal{B}, \mu)$ be a prob. space.

for $A \in \mathcal{B}$, $\mu(A) > 0$ we define: $\mu|_A(B) = \frac{\mu(A \cap B)}{\mu(A)}$

This is the (normalized) restriction of μ to A.

Note: some texts denote: $\mu|_A(B) = \mu(A \cap B)$. we will denote $\mu||_A(B) = \mu(A \cap B)$

if ξ is a partition $\xi = (A_1, A_2, \dots)$ then

$$\mu = \sum_i \mu||_{A_i} = \sum_i \mu(A_i)\, \mu|_{A_i}$$

for any f integrable $\int f\, d\mu = \sum_i \mu(A_i) \int f\, d\mu|_{A_i}$

for partitions ξ, η define information:

$$I_\mu(\xi|\eta)(x) = -\log\left(\frac{\mu([x]_{\xi \vee \eta})}{\mu([x]_\eta)}\right) = I_{\mu|_{[x]_\eta}}(\xi)(x)$$

$H_\mu(\xi|\eta) = \sum_j \mu(B_j) H_{\mu/B_j}(\xi)$ אנטרופיה תלויה:

Then:

1) $H_\mu(\xi|\eta) = \int I_\mu(\xi|\eta) d\mu$ \ $H_\mu(\xi) = \int I_\mu(\xi) d\mu$

2) additivity: $\to I_\mu(\xi \vee \eta) = I_\mu(\eta) + I_\mu(\xi|\eta)$

$\to H_\mu(\xi \vee \eta) = H_\mu(\eta) + H_\mu(\xi|\eta)$

In particular, if $H_\mu(\eta) < \infty$ then $H_\mu(\xi|\eta) = H_\mu(\xi \vee \eta) - H_\mu(\eta)$

Note: (2) is (4) from previous lecture.

$H(p_1, p_2, \dots) = -\sum p_i \log(p_i)$

Also if $\xi < \eta$, $H_\mu(\xi) \geq H_\mu(\eta)$

3) $H_\mu(\xi \vee \eta) \leq H_\mu(\xi) + H_\mu(\eta)$

3') $H_\mu(\xi \vee \eta | \alpha) \leq H_\mu(\xi|\alpha) + H_\mu(\eta|\alpha)$

4) If $H_\mu(\alpha) < \infty$, $H_\mu(\eta) < \infty$ then:

$H_\mu(\xi | \eta \vee \alpha) \leq H_\mu(\xi|\alpha)$

(in particular, if $\eta < \alpha$ then $H_\mu(\xi|\eta) \leq H_\mu(\xi|\alpha)$)

Proofs:

1) easy to check

2) By integration, it is enough to check additivity of I_μ.

$I_\mu(\xi \vee \eta) = -\log(\mu([x]_{\xi \vee \eta})) = -\log(\mu([x]_\xi \cap [x]_\eta)) =$

$= -\log(\mu([x]_\eta)) - \log\left(\frac{\mu([x]_\xi \cap [x]_\eta)}{\mu([x]_\eta)}\right) = I_\mu(\eta)(x) + I_\mu(\xi|\eta)(x)$

3) $H_\mu(\xi|\eta) = \sum_j \mu(B_j) \cdot -\sum_i \phi\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) = -\sum_i \sum_j \mu(B_j) \phi\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) \leq$

$[\phi(x) = x \log x]$ Jensens inequality

$\leq -\sum_i \phi\left(\sum_j \mu(B_j) \frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) = -\sum_i \phi(\mu(A_i)) = H_\mu(\xi)$

now add $H_\mu(\eta)$ to both sides & use (2)

4) $\alpha = \{ C_k \mid k \in \mathcal{R} \}$

$$H_\mu(\xi \mid \eta \vee \alpha) = H_\mu(\xi \vee \eta \vee \alpha) - H_\mu(\eta \vee \alpha) =$$

$$= H_\mu(\alpha) + H_\mu(\xi \vee \eta \mid \alpha) - H_\mu(\alpha) - H_\mu(\eta \mid \alpha) =$$

$$= \sum_k \mu(C_k) \left(H_{\mu|C_k}(\xi \vee \eta) - H_{\mu|C_k}(\eta) \right) = \sum_k \mu(C_k)$$

Entropy of P.P.S

A P.P.S (probability preserving system) is $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a probability space, $T: X \to X$, $T^{-1}\mathcal{B} \subset \mathcal{B}$ s.t. $\forall A \in \mathcal{B}$. $\mu(A) = \mu(T^{-1}(A))$.

Lemma:

if ξ is a partition, $T^{-1}\xi = \{ T^{-1}A \mid A \in \xi \}$. Then for any two partitions ξ, η, $H_\mu(\xi) = H_\mu(T^{-1}\xi)$, $I_\mu(\xi)(Tx) = I_\mu(T^{-1}\xi)(x)$

$H_\mu(\xi|\eta) = H_\mu(T^{-1}\xi \mid T^{-1}\eta)$, $I_\mu(\xi|\eta)(Tx) = I_\mu(T^{-1}\xi \mid T^{-1}\eta)(x)$

Proof:

Only need to show $I_\mu(\xi|\eta)(Tx) = I_\mu(T^{-1}\xi \mid T^{-1}\eta)(x)$.

Note: $T^{-1}([Tx]_\eta) = [x]_{T^{-1}\eta}$

Thus $I_\mu(\xi|\eta)(Tx) = -\log \dfrac{\mu([Tx]_\xi \cap [Tx]_\eta)}{\mu([Tx]_\eta)} =$

$$= -\log \frac{\mu([x]_{T^{-1}\xi} \cap [x]_{T^{-1}\eta})}{\mu([x]_{T^{-1}\eta})} = I_\mu(T^{-1}\xi \mid T^{-1}\eta)(x)$$

T^{-1} preserves μ
& preserves intesections

Fecete's Lemma:

Suppose (a_n) is a sequence of real numbers, which is sub-additive, i.e. $a_{n+m} \le a_n + a_m$ $\forall n, m$.

Then $\lim\limits_{n \to \infty} \frac{1}{n} a_n = \inf\limits_n \frac{1}{n} a_n$ (in particular, limit exists).

we will prove the lemma & deduce:

Cor.:

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

This number will be called μ-entropy of T w.r.t ξ & denoted $h_\mu(T, \xi)$

Proof:

let $a = \inf \frac{1}{n} a_n$. Assume $a > -\infty$ (we will work with $a_n \geq 0$). let $\varepsilon > 0$, choose k s.t $\frac{a_k}{k} < a + \varepsilon$.

For each n large, write $n = mk + j$, $0 \leq j < k$

($m = m_n$, $j = j_n$, j_n is bounded, $m_n \xrightarrow[n\to\infty]{} \infty$).

Then $$\frac{a_n}{n} \leq \frac{a_{mk}}{j+mk} + \frac{a_j}{j+mk} \leq \frac{a_{mk}}{mk} + \frac{a_j}{mk} \leq$$

$$\leq \frac{m a_k}{mk} + \frac{j a_1}{mk} \leq a + \varepsilon + \frac{j a_1}{mk}$$

$\frac{j a_1}{mk} \xrightarrow{n\to\infty} 0$

So for all n large enough, $\frac{a_n}{n} < a + 2\varepsilon$

$$a \leq \liminf \frac{a_n}{n} \leq \limsup \frac{a_n}{n} \leq a$$

$$\Longrightarrow \lim_{n\to\infty} \frac{a_n}{n} = a$$

Cor:

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

Proof:

Define $a_n = H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right)$, need to show $a_{n+m} \leq a_n + a_m$.

$$a_n + a_m = H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + H_\mu\left(\bigvee_{i=0}^{m-1} T^{-i}\xi\right) =$$

Lemma apply T^{-n}

$$= H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + H_\mu\left(\bigvee_{j=n}^{n+m-1} T^{-j}(\xi)\right) \geq H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi \cup \bigvee_{j=n}^{n+m-1} T^{-j}\xi\right) =$$

$$= a_{n+m}$$

ξ - experiment
T - time

Def:

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

intuition: this is the average amount of expected information from repeating experiment ξ n times as $n \to \infty$

$$h_\mu(T) = \sup_{H_\mu(\xi)<\infty} h_\mu(T,\xi)$$

this is the amount of information you can get from repeating the optimal experiment

Examples:

(o) $T = Id$, μ is any measure, ξ any partition, $T^{-1}\xi = \xi$

hence $\bigvee_{i=0}^{n-1} T^{-i}\xi = \xi$ $\quad H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) = H_\mu(\xi) < \infty$

Dividing by n gives $h_\mu(T, \xi) = 0$ whenever $H_\mu(\xi) < \infty$

$\Longrightarrow H_\mu(Id) = 0$.

(Exercise: $\exists p_1, p_2, \dots$ s.t $H(p_1, p_2, \dots) = \infty$, where $H_\mu(\vec{p}) = -\sum p_i \log p_i$)

Properties: (entropy conditioned on future)

$(X, \mathcal{B}, \mu, T)$ a p.p.s, ξ partition, $H_\mu(\xi) < \infty$

Then: $h_\mu(T, \xi) = \lim_{n \to \infty} H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{n} T^{-i}\xi\right)$

Intuitive meaning:

This is the average amount of information obtained by an experiment given all future experiments.

Recall: $a_n \xrightarrow[n\to\infty]{} \alpha$ then $\frac{1}{n}\sum_{i=0}^{n-1} a_i \longrightarrow \alpha$ (the Cesaro average)

Proof of property:

By monotoniaty, $H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{n} T^{-i}\xi\right)$ is decreasing, so limit of RHS exists, denote it by α.

use additivity, & induction, to get:

$$H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) = H_\mu\left(\bigvee_{i=0}^{n-2} T^{-i}\xi \,\middle|\, T^{-(n-1)}\xi\right) + H_\mu\left(T^{-(n-1)}\xi\right) = \dots =$$

$$= H_\mu\left(T^{-(n-1)}\xi\right) + H_\mu\left(T^{-(n-2)}\xi \,\middle|\, T^{-(n-1)}\xi\right) + \dots + H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{n-1} T^{-i}\xi\right) =$$

by invariance

$$= \underbrace{H_\mu(\xi) + H_\mu(\xi \mid T^{-1}\xi) + \dots + H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{n-1} T^{-i}\xi\right)}_{(*)}$$

so $\frac{1}{n}\left(H_\mu \xi \,\middle|\, \bigvee_{i=0}^{n-1} T^{-i}\xi\right) = \frac{1}{n}\cdot(*) \underset{\uparrow}{=} \alpha$

a cesaro average of a sequence tending to α.

Examples (Cont.)

1) let $X = \{0,1\}^{\mathbb{Z}}$, $\mathcal{B}$ - Borel σ-algebra w.r.t Tychonov topology $= \sigma\left(\left\{C(b_j)_{k \le j \le l} \;\middle|\; \forall k \le l \in \mathbb{Z}\ \forall j = k, \dots, l\ b_j = \{0,1\}\right\}\right)$

$\Big[$ where $C(b_j)_{k \le j \le l} = \left\{ (a_i) \in X \;\middle|\; \forall j = k, k+1, \dots, l \quad a_j = b_j \right\}$

$C((0)_{j=0}) = \left\{ (a_i) \;\middle|\; a_0 = 0 \right\}$, $C((1)_{j=0}) = \left\{ (a_i) \;\middle|\; a_0 = 1 \right\}$ $\Big]$

$\mu = \underbrace{\left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right)}_{\text{measure on } \{0,1\}}^{\otimes \mathbb{Z}}$ = "coin tossing measure"

Bernoulli measure for prob's $\frac{1}{2}, \frac{1}{2}$.

μ is determined (via Caratheodory extension theorem) by the requirement:

$$\mu\left(C(b_j)_{k \le i \le l}\right) = \left(\frac{1}{2}\right)^{l-k+1}$$

$\sigma : X \longrightarrow X$ is the left shift:

$(\sigma((a_i)))_i = a_{i+1}$

$a_i = \quad \dots, 1, 0, 0, 1, 0, 0, 1, 1, 0, \dots$

$\sigma(a_i) = \quad \dots, 1, 0, 0, 1, 0, 0, 1, 1, 0, \dots$ (moving time forward)

let $\xi = \{A_0, A_1\}$, $A_0 = C((0)_{j=0})$, $A_1 = C((1)_{j=0})$

$H_\mu(\xi) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = \log 2$

$$\bigvee_{i=0}^{n-1} \sigma^{-i}(\xi) = \left\{ C(b_j)_{0 \le j \le n-1} \;\middle|\; b_j \in \{0,1\},\ 0 \le j \le n-1 \right\}$$

a partition into 2^n sets, each of measure 2^{-n}.

$$H_\mu\left(\bigvee_{i=0}^{n-1} \sigma^{-i}(\xi)\right) = -\left(\sum_{A \in \bigvee_0^{n-1} T^{-i}\xi} \mu(A)\log(\mu(A))\right) = -2^n\left(2^{-n}\log 2^{-n}\right) =$$

$= n\log 2$

constant - coin tosses are independent
you get no information of the future by knowing the past

$$\frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} \sigma^{-i}(\xi)\right) = \log 2 \longrightarrow \log 2 = h_\mu(\sigma, \xi)$$

$H_\mu(\xi \vee \eta) = H_\mu(\xi) + H_\mu(\eta) \iff$ for all i, j $\mu(A_i \cap A_j) = \mu(A_i) \cdot \mu(B_j)$
(Exercise)

(in this case we say η, ξ are independent)
assume that $H_\mu(\xi), H_\mu(\eta) < \infty$

<u>Basic Properties</u>: $(X, \mathcal{B}, \mu, T)$ p.p.s. ξ, η partitions of finite entropy

Then: (1). (trivial bound) $h_\mu(T, \xi) \leq H_\mu(\xi)$

$$h_\mu(T,\xi) = \inf \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right)$$

(2) (sub additivity) $h_\mu(T, \xi \vee \eta) \leq h_\mu(T, \xi) + h_\mu(T, \eta)$

(3) (Continuity bound) $h_\mu(T, \eta) \leq h_\mu(T, \xi) + H_\mu(\eta \mid \xi)$

<u>Proof of (2)</u>:

For each n, $H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi \vee \eta)\right) = H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi \vee \bigvee_{i=0}^{n-1} T^{-i}\eta\right) \leq$

$$\underset{\text{monotonicity}}{\leq} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\eta\right)$$

divide both sides by n & let $n \to \infty$

<u>Proof of (3)</u>:

$$h_\mu(T, \eta) = \lim_{n \to \infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\eta\right) \underset{(2)}{\leq} \lim_{n \to \infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi \vee \eta)\right) =$$

$$= \lim_{n \to \infty} \left[\frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\eta) \,\middle|\, T^{-i}\xi\right)\right] \underset{(3),(4)}{\leq}$$

$$\leq h_\mu(T, \xi) + \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} H_\mu\left(T^{-i}\eta \mid T^{-i}\xi\right) \underset{\text{invariance}}{=}$$

$$= h_\mu(T, \xi) + H_\mu(\eta \mid \xi)$$