

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 $x \in A_i$
- $I_\mu(\xi) : X \rightarrow [0, 1]$
 $I_\mu(\xi)(x) = -\log(\mu([x]_\xi))$
- $H_\mu(\xi) = \int I_\mu(\xi) d\mu = -\sum_{i \in I} \mu(A_i) \log(\mu(A_i))$

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:

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

$$H_\mu(\xi|\eta) = \sum_j \mu(B_j) H_{\mu|B_j}(\xi) \quad \text{; יחסיות תנאים}$$

Then:

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

$$2) \text{ additivity: } \rightarrow I_\mu(\xi \vee \eta) = I_\mu(\eta) + I_\mu(\xi|\eta) \\ \rightarrow H_\mu(\xi \vee \eta) = H_\mu(\eta) + H_\mu(\xi|\eta)$$

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 .

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

$$3) H_\mu(\xi|\eta) = \sum_j \mu(B_j) \cdot \sum_i \phi\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) = - \sum_i \sum_j \mu(B_j) \phi\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) \leq \\ \left[\phi(x) = x \log x \right] \quad \text{Jensen's inequality}$$

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

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

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

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

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

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

Entropy of P.P.S

A P.P.S (Probability Preserving System) is (X, B, μ, T)

where (X, B, μ) is a probability space, $T: X \rightarrow X$, $T^{-1}B \subset B$

s.t. $\forall A \in B$. $\mu(A) = \mu(T^{-1}(A))$

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

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

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

T^{-1} preserves μ
& preserves intersections

Fekete's Lemma:

Suppose (a_n) is a sequence of real numbers, which is

sub-additive, i.e. $a_{n+m} \leq a_n + a_m \quad \forall n, m$.

Then $\lim_{n \rightarrow \infty} \frac{1}{n} a_n = \inf_n \frac{1}{n} a_n$ (in particular, limit exists).

we will prove the lemma & deduce:

Cor:

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

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 $\epsilon > 0$, choose K s.t. $\frac{a_K}{K} < a + \epsilon$. For each n large, write $n = mK + j$, $0 \leq j < K$ ($m = m_n$, $j = j_n$, j_n is bounded, $m_n \xrightarrow{n \rightarrow \infty} \infty$).

$$\begin{aligned} \text{Then } \frac{a_n}{n} &\leq \frac{a_{mK}}{j+mK} + \frac{a_j}{j+mK} \leq \frac{a_{mK}}{mK} + \frac{a_j}{mK} \leq \\ &\leq \frac{ma_K}{mK} + \frac{j a_1}{mK} \leq a + \epsilon + \frac{j a_1}{mK} \end{aligned}$$

$\xrightarrow[n \rightarrow \infty]{0}$

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

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

$$\Rightarrow \lim_{n \rightarrow \infty} \frac{a_n}{n} = a$$

Cor:

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

Proof:

Define $a_n = H_\mu \left(\bigvee_{i=0}^{n-1} T^{-i} \zeta \right)$, need to show

$$a_{n+m} \leq a_n + a_m.$$

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

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

Lemma
apply T^{-n}

$$= a_{n+m}$$

Def:

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

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

ζ - experiment
 T - time

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

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

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

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

Intuitive meaning:

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

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

Proof of property:

By monotonicity, $H_\mu\left(\xi \mid \bigvee_{i=1}^n T^{-i}\xi\right)$ is decreasing. So limit of RHS exists, denote it by $\mathbb{Z}$.

use additivity, & induction, to get:

$$\begin{aligned} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) &= H_\mu\left(\bigvee_{i=0}^{n-2} T^{-i}\xi \mid T^{-(n-1)}\xi\right) + H_\mu(T^{-(n-1)}\xi) = \dots = \\ &= H_\mu(T^{-(n-1)}\xi) + H_\mu(T^{-(n-2)}\xi \mid T^{-(n-1)}\xi) + \dots + H_\mu\left(\xi \mid \bigvee_{i=1}^{n-1} T^{-i}\xi\right) = \end{aligned}$$

$$\stackrel{\text{by invariance}}{\rightarrow} = H_\mu(\xi) + H_\mu(\xi \mid T^{-1}\xi) + \dots + H_\mu\left(\xi \mid \bigvee_{i=1}^{n-1} T^{-i}\xi\right)$$

$$\text{so } \frac{1}{n} \left(H_\mu\left(\xi \mid \bigvee_{i=0}^{n-1} T^{-i}\xi\right) \right) = \frac{1}{n} \cdot (n) = \mathbb{Z}$$

a cesaro average of a sequence tending to $\mathbb{Z}$.

Example 5 (Cont.)

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

$$\left[\begin{array}{l} \text{where } C(b_j)_{k \leq j \leq l} = \left\{ (a_i) \in X \mid \forall j = k, k+1, \dots, l \quad a_j = b_j \right\} \\ C((0)_{j=0}) = \left\{ (a_i) \mid a_0 = 0 \right\}, \quad C((1)_{j=0}) = \left\{ (a_i) \mid a_0 = 1 \right\} \end{array} \right]$$

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

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

μ is determined (via Carathéodory extension theorem) by the requirements:

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

$\tau: X \rightarrow X$ is the left shift:

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

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

$$\tau(a_i) = \dots, 1, 0, 0, 1, 0, 1, 1, 0, \dots \quad \left(\begin{array}{c} \text{moving time} \\ \text{forward} \end{array} \right)$$

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

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

$$\bigvee_{i=0}^{n-1} \tau^{-i}(\xi) = \left\{ C(b_j)_{0 \leq j \leq n-1} \mid b_j \in \{0,1\}, 0 \leq j \leq n-1 \right\}$$

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

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

$$= n \log 2$$

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

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

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

(in this case we say ξ, η are independent)

assume that $H_\mu(\xi), H_\mu(\eta) < \infty$

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

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

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

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

Proof of (2):

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

monotonicity

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

Proof of (3):

$$h_\mu(T, \eta) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i} \eta\right) \stackrel{(2)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi \vee \eta)\right) = \\ = \lim_{n \rightarrow \infty} \left[\frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i} \xi\right) + \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\eta) \mid \bigvee_{i=0}^{n-1} T^{-i} \xi\right) \right] \stackrel{(3), (4)}{\leq} \\ \leq h_\mu(T, \xi) + \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} H_\mu(T^{-i} \eta \mid T^{-i} \xi) \stackrel{\text{invariance}}{=} \\ = h_\mu(T, \xi) + H_\mu(\eta | \xi)$$

□