(NewChap8) 6 Entropy – History

Entropy

Clausius 1847
macroscopic entropy
in Thermodynamics

J.W. Gibbs ~~1902~~ 1873-78

1916-2001 Claude Shannon 1948
Information theory
math theory of communication

Kolmogorov – Sinai ++
1958 1959

metric entropy
measure theoretic entropy

ex upshot: $x \mapsto 2x \bmod 1$ and $x \mapsto 3x \bmod 1$ are not measure isomorphic

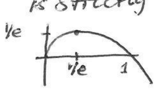
Def 6.1 Let $(X, \mathcal{B}_X, \mu)$ be a probability space and $\mathcal{A}$ a finite sub-algebra of $\mathcal{B}_X$ generated by a partition $\xi_{\mathcal{A}} = \{A_1, \dots, A_k\}$ $k < \infty$

The entropy of $\mathcal{A}$ wrt μ is

$$H(\mathcal{A}) = H_\mu(\mathcal{A}) = H(\xi_A) = -\sum_{i=1}^{k} \mu(A_i) \log \mu(A_i) \geq 0$$

Intuition: H measures the average amount of uncertainty removed by (or information gained) by performing a probabilistic experiment

Lemma 6.2 The function $\phi(t) = -t\log t$ is strictly concave on $t \geq 0$.



By Jensen's inequality for $x_i \geq 0$, $p_i \geq 0$, $\sum p_i = 1$

$$\boxed{\phi(\sum p_i x_i) \geq \sum p_i \phi(x_i)} \quad (*)$$

with equality iff for every i with $p_i > 0$ the x_i's are equal.

$$H(\mathcal{A}) = \sum_{i=1}^{k} \phi(\mu(A_i))$$

Proposition 6.3
If $\xi_A = \{A_1, \dots, A_k\}$ then
$H(\mathcal{A}) \leq \log k$
with equality iff $\mu(A_i) = \frac{1}{k}$ all i.

proof: $x_i = \mu(A_i)$ $p_i = \frac{1}{k}$ yields

$$\sum p_i \phi(x_i) \leq \phi(\sum p_i x_i) = \phi(\tfrac{1}{k})$$

$$\tfrac{1}{k} \sum \phi(x_i) = \tfrac{1}{k} H(\mathcal{A}) \leq \tfrac{1}{k} \log k$$

Ex : A dice  $X = \{1, \dots, 6\}$ equi-prob

$$H(\mathcal{B}_X) = \sum_{1}^{6} \tfrac{1}{6} \log 6 = \log 6$$

$\xi_{\mathcal{A}} = \{\text{even}, \text{odd}\} = \{(1,3,5), (2,4,6)\}$

$H(\mathcal{A}) = \log 2$

$\xi_{\mathcal{C}} = \{(1), (2,3), (4,5,6)\}$

$$H(\mathcal{C}) = \tfrac{1}{6} \log(6 \cdot 3^2 \cdot 2^3) < \log 3$$

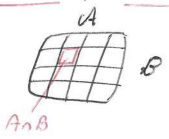
$$= \tfrac{1}{6} \log 6 + \tfrac{1}{3} \log 3 + \tfrac{1}{2} \log 2$$

Def 6.4 Given partitions $\mathcal{A}, \mathcal{B}$ we define the join of $\mathcal{A}$ and $\mathcal{B}$

$$\mathcal{A} \vee \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$$

which is thus a refinement of both $\mathcal{A}$ and $\mathcal{B}$

(i.e. every $A \in \mathcal{A}$ is a disjoint union of elements in $\mathcal{A} \vee \mathcal{B}$)



Def 6.5 We define the conditional entropy of $\mathcal{A}$ given $\mathcal{B}$:

$$H(\mathcal{A}|\mathcal{B}) = -\sum_{\substack{A \in \mathcal{A} \\ B \in \mathcal{B}}} P(A \cap B) \log P(A|B)$$

$$= \sum_{A, B} P(B)\ \Phi(P(A|B)) \geq 0$$

$P(B)\,P(A|B)$

Prop 6.6 For finite partitions $\mathcal{A}$ and $\mathcal{B}$ we have:

$$H(\mathcal{A}|\mathcal{B}) = H(\mathcal{A} \vee \mathcal{B}) - H(\mathcal{B})$$

proof: $H(\mathcal{A}|\mathcal{B}) =$

$$- \sum P(A \cap B) \log \frac{P(A \cap B)}{P(B)} =$$

$$- \sum P(A \cap B) \log P(A \cap B) + \sum P(A \cap B) \log P(B)$$

$$= H(\mathcal{A} \vee \mathcal{B}) + \sum P(B) \log P(B)$$

$$= H(\mathcal{A} \vee \mathcal{B}) - H(\mathcal{B})$$

Coroll 6.7 We have:

$$H(\mathcal{A} \vee \mathcal{B} | \mathcal{B}) = H(\mathcal{A}|\mathcal{B})$$

$$H(\mathcal{B}|\mathcal{B}) = 0$$

since $\mathcal{A} \vee \mathcal{B} \vee \mathcal{B} = \mathcal{A} \vee \mathcal{B}$

$\mathcal{B} \vee \mathcal{B} = \mathcal{B}$

Examples: 6.8

Independent partitions:
Suppose $P(A \cap B) = P(A)P(B)$ or for all A, B with $P(B) > 0$ $P(A|B) = P(A)$

Then

$$H(\mathcal{A}|\mathcal{B}) = \sum P(B)\ \Phi(P(A|B))$$
$$= \sum P(B)\ \Phi(P(A))$$
$$= \sum \Phi(P(A))$$
$$= H(\mathcal{A})$$

when $\mathcal{A}$ and $\mathcal{B}$ are indep^s knowing $\mathcal{B}$ yields no information as regards $\mathcal{A}$.

Refinements. Suppose $P(A|B) = 0$ or 1 all A, B

$$H(\mathcal{A}|\mathcal{B}) = -\sum P(B)\ \Phi(P(A|B))$$
$$= 0$$

Here $\mathcal{B}$ is a refinement of $\mathcal{A}$.

Knowing $\mathcal{B}$ we know for sure the result in $\mathcal{A}$.

Dice $X = \{1, \dots, 6\}$

$\mathcal{B}_*$ the power set partition

$\mathcal{A} = \{(2,4,6), (1,3,5)\}$ even odd

$$H(\mathcal{B}_*|\mathcal{A}) = H(\mathcal{B}) - H(\mathcal{A})$$
$$= \log 6 - \log 2$$
$$= \log 3.$$

3 choices left to decide

Prop 6.9

For finite partitions $\mathcal{A}, \mathcal{B}, \mathcal{C}$:

$H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{B}|\mathcal{A} \vee \mathcal{C})$

proof:

$H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) =$

$H(\mathcal{A} \vee \mathcal{B} \vee \mathcal{C}) - H(\mathcal{C}) =$

$H(\mathcal{A} \vee \mathcal{B} \vee \mathcal{C}) - H(\mathcal{A} \vee \mathcal{C}) +$

$H(\mathcal{A} \vee \mathcal{C}) - H(\mathcal{C}) =$

$H(\mathcal{B}|\mathcal{A} \vee \mathcal{C}) + H(\mathcal{A}|\mathcal{C})$ //

mod 0:

Def 6.10 Given finite meas partitions $\mathcal{A}$, $\mathcal{C}$ we say that

$\mathcal{A} \overset{\circ}{\subset} \mathcal{C}$

provided every $A \in \mathcal{A}$ is a finite disjoint union of elements in $\mathcal{C}$ (mod 0)

Thm 6.11 Given finite $\mathcal{A}, \mathcal{B}$ subalgebras of $\mathcal{B}_X$ we have:

1) $H(\mathcal{A}|\mathcal{C}) = 0 \Leftrightarrow \mathcal{A} \overset{\circ}{\subset} \mathcal{C}$

2) $H(\mathcal{A}|\mathcal{C}) = H(\mathcal{A})$ iff $\mathcal{A}$ and $\mathcal{C}$ are independent: $\mu(A \cap C) = \mu(A)\mu(C)$ $\forall A \in \mathcal{A}, C \in \mathcal{C}$

later

proof: $H(\mathcal{A}|\mathcal{C}) = 0$ iff for every A, C in the partitions

$\mu(A \cap C) \cdot \log \frac{\mu(A \cap C)}{\mu(C)} = 0$

which happens if $\mu(A \cap C) = 0$ or $\mu(A \cap C) = \mu(C)$ so $C \subset A$ (mod 0).

In example 6.8 we showed $\mathcal{A}, \mathcal{B}$ indep $\Rightarrow$ $H(\mathcal{A}|\mathcal{B}) = H(\mathcal{A})$

Conversely for fixed partition element A:

$\sum_{C \in \mathcal{C}} \mu(C)\, \phi(\mu(A|C)) \leq \phi(\mu(A))$

by Jensen's inequality with equality iff $\mu(A|C) = \mu(A)$ whenever $\mu(C) > 0$.

i.e. $\mu(A \cap C) = \mu(A)\mu(C)$ for every partition elements A, C whence also for disjoint union.

~~Thm 6.12~~

$d(\mathcal{A}, \mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{C}|\mathcal{A})$ defines a metric on the space of finite subalgebras of $\mathcal{B}_X$ (mod 0)

proof: $d(\mathcal{A}, \mathcal{C}) \geq 0$ with equality iff $\mathcal{A} \overset{\circ}{\subset} \mathcal{C}$ and $\mathcal{C} \overset{\circ}{\subset} \mathcal{A}$.

Also

$H(\mathcal{A}|\mathcal{C}) \leq H(\mathcal{A} \vee \mathcal{B}|\mathcal{C})$

$= H(\mathcal{B}|\mathcal{C}) + H(\mathcal{A}|\mathcal{B} \vee \mathcal{C})$

$\leq H(\mathcal{B}|\mathcal{C}) + H(\mathcal{A}|\mathcal{B})$

gives the triangle inequality

Combined conditional probability

$$P(A\cap B\cap C) = \frac{P(A\cap B\cap C)}{P(B\cap C)}\,\frac{P(B\cap C)}{P(C)}\,P(C)$$

$$= P(A|B\cap C)\,P(B|C)\,P(C)$$

$\sum_{B\in\mathcal{B}}$: $\quad \frac{P(A\cap C)}{P(C)} = \sum_{B\in\mathcal{B}} P(A|B\cap C)\,\underline{P(B|C)}$ $\quad \sum_{B} P(B|C)=1$

A,C fixed

$\phi(t) = -t\log t$

$$\phi(P(A|C)) \ge \sum_B \phi(P(A|B\cap C))\,P(B|C)$$

$$P(C)\,\phi(P(A|C)) \ge \sum_B P(B\cap C)\,\phi(P(A|B\cap C))$$

Sum over partitions $\mathcal{A}, \mathcal{C}$:

$$\sum_{A,C} P(C)\,\phi(P(A|C)) \ge \sum_{A,B,C} P(B\cap C)\,\phi(P(A|B\cap C))$$

or $\quad H(\mathcal{A}|\mathcal{C}) \ge H(\mathcal{A}|\mathcal{B}\vee\mathcal{C})$

put up before

we have shown:

Proposition 6.12
Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be finite partitions.
Then
$$H(\mathcal{A}|\mathcal{C}) \ge H(\mathcal{A}|\mathcal{B}\vee\mathcal{C})$$

knowing the outcome of $\mathcal{B}$ and $\mathcal{C}$ leaves less uncertainty for $\mathcal{A}$ than just knowing $\mathcal{C}$.

The more information we know a priori the less we gain by performing an experiment

exs $\quad$ for any $\mathcal{C}$

$$H(\mathcal{A}) = H(\mathcal{A}|\{\emptyset,X\}) \ge H(\mathcal{A}|\mathcal{C})$$

$$0 \le H(\mathcal{A}|\mathcal{B}_X) \le H(\mathcal{A}|\mathcal{A}) = 0$$

Corollary 6.13 I
Let $\mathcal{A}, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \dots$ be finite partitions with
$$\mathcal{C}_0 \subset \mathcal{C}_1 \subset \mathcal{C}_2 \subset \dots$$
Then
$$H(\mathcal{A}|\mathcal{C}_0) \ge H(\mathcal{A}|\mathcal{C}_1) \ge \dots \quad (30)$$

follows from $\mathcal{C}_{n+1} = \mathcal{C}_n \vee \mathcal{C}_{n+1}$

Corollary 6.13 II
Entropy is subadditive

$$H(\mathcal{A}\vee\mathcal{B}) \le H(\mathcal{A}) + H(\mathcal{B})$$
$$H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) \le H(\mathcal{A}|\mathcal{C}) + H(\mathcal{B}|\mathcal{C})$$

proof: just use
$H(\mathcal{B}|\mathcal{A}\vee\mathcal{C}) \le H(\mathcal{B}|\mathcal{C})$
in Prop 6.9.

Theorem 6.14
$d(\mathcal{A}|\mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{C}|\mathcal{A})$ defines a metric on the space of finite sub-σ-alg of $\mathcal{B}_X$ (mod 0).

$d(\mathcal{A}|\mathcal{C}) = 0$ iff $\mathcal{A} \doteq \mathcal{C}$.

By thm 6.11 $\quad d(\mathcal{A}|\mathcal{C}) = 0$
iff $\quad H(\mathcal{A}|\mathcal{C}) = H(\mathcal{C}|\mathcal{A}) = 0$
iff $\quad \mathcal{A} \dot\subset \mathcal{C}$ and $\mathcal{C} \dot\subset \mathcal{A}$
$\Leftrightarrow \quad \mathcal{A} \doteq \mathcal{C}$

Triangle ineq:

$$H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + \overbrace{H(\mathcal{B}|\mathcal{A}\vee\mathcal{C})}^{\ge 0}$$
$$= H(\mathcal{B}|\mathcal{C}) + H(\mathcal{A}|\mathcal{B}\vee\mathcal{C})$$
$$\le H(\mathcal{B}|\mathcal{C}) + H(\mathcal{A}|\mathcal{B})$$

$\Rightarrow H(\mathcal{A}|\mathcal{C}) \le H(\mathcal{A}|\mathcal{B}) + H(\mathcal{B}|\mathcal{C})$

by symmetry
$H(\mathcal{C}|\mathcal{A}) \le H(\mathcal{C}|\mathcal{B}) + H(\mathcal{B}|\mathcal{A})$

$\Rightarrow d(\mathcal{A},\mathcal{C}) \le d(\mathcal{A},\mathcal{B}) + d(\mathcal{B},\mathcal{C})$

Entropy of a meas. preserving transf.

Let $(X, \mathcal{B}_X, \mu, T)$ be a meas pres transf of a proba space.

Lemma 6.15: Given finite meas partitions $\mathcal{A}$ and $\mathcal{B}$:

$$H(T^{-1}\mathcal{A}) = H(\mathcal{A})$$
$$H(T^{-1}\mathcal{A} \mid T^{-1}\mathcal{B}) = H(\mathcal{A} \mid \mathcal{B})$$

proof:
When $\xi_{\mathcal{A}} = \{A_1, \dots, A_k\}$ is the $\mathcal{A}$-partition

then $\xi_{T^{-1}\mathcal{A}} = \{T^{-1}A_1, \dots, T^{-1}A_k\}$ is the $T^{-1}\mathcal{A}$ partition and they have the same probabilities

Same for $\mathcal{A} \vee \mathcal{B}$ for which
$T^{-1}(\mathcal{A} \vee \mathcal{B}) = T^{-1}\mathcal{A} \vee T^{-1}\mathcal{B}$ //

We may use T to dynamically increase the entropy of a partition

Given $\xi_{\mathcal{A}} = \{A_1, \dots, A_n\}$ we set for $p \geq 1$

$$\mathcal{A}_p = \mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(p-1)}\mathcal{A} = \bigvee_{i=0}^{p-1} T^{-i}\mathcal{A}$$
$$\mathcal{A}_0 \equiv \{\emptyset, X\}$$

Def 6.16 We set

$$h_\mu(T, \mathcal{A}) = h(T, \mathcal{A}) = \lim_{n \to \infty} \frac{1}{p} H_\mu(\mathcal{A}_p)$$

It is called the measure theoretic entropy of the map T wrt the partition $\mathcal{A}$ and the meas μ.

$$h_\mu(T) = \sup_{\text{finite } \mathcal{A}} h_\mu(T, \mathcal{A}) \in [0, +\infty]$$

is the measure μ-entropy of T.

Theorem 6.17

$$h_\mu(T,\mathcal{A}) = \lim_n \frac{1}{n} H_\mu(\mathcal{A}_n) = \inf_{n\geq 0} \frac{1}{n} H_\mu(\mathcal{A}_n) = \lim_n H_\mu(\mathcal{A} / T^{-1}\mathcal{A}_n)$$

(so in particular $\leq H_\mu(\mathcal{A})$)

---

proof: We have

$$\underline{\mathcal{A}_n} = \bigvee_0^{n-1} T^{-i}\mathcal{A} = \mathcal{A} \vee T^{-1} \bigvee_0^{n-2} T^{-i}\mathcal{A} = \underline{\mathcal{A} \vee T^{-1}\mathcal{A}_{n-1}}$$

By corollary 6.13:

$h_1 = H(\mathcal{A})$, $h_k = H(\mathcal{A} / T^{-1}\mathcal{A}_{k-1})$, $k \geq 2$

is a non-increasing (positive) Sequence so has a limit

$$h_0 \geq h_\infty = \lim_k h_k \geq 0$$

$$\begin{aligned} H(\mathcal{A}_n) &= H(\mathcal{A}_n / T^{-1}\mathcal{A}_{n-1}) + H(T^{-1}\mathcal{A}_{n-1}) \\ &= H(\mathcal{A} / T^{-1}\mathcal{A}_{n-1}) + H(\mathcal{A}_{n-1}) \\ &= h_n + H(\mathcal{A}_{n-1}) \\ &= \dots = \sum_{k=0}^{n-1} h_k = h_{n} + \dots + h_1 \quad h_1 = \end{aligned}$$

$$\Rightarrow \frac{1}{n} H(\mathcal{A}_n) \searrow_n h_\infty = h(T,\mathcal{A})$$

Entropy inequality
and conjugacy invariance.

6.18 $\mathfrak{X}_1 =$
Theorem: Let $(X_1, \mathcal{B}_1, m_1, T_1)$ and
$\mathfrak{X}_2 = (X_2, \mathcal{B}_2, m_2, T_2)$ be measure pres. transformations of proba spaces.

a) Suppose that $\mathfrak{X}_2$ is a factor of $\mathfrak{X}_1$. Then
$$h(\mathfrak{X}_2) \le h(\mathfrak{X}_1)$$
b) If $\mathfrak{X}_1$ and $\mathfrak{X}_2$ are isomorphic then
$$h(\mathfrak{X}_2) = h(\mathfrak{X}_1)$$

proof: Let $\phi : X_1 \to X_2$ be a meas. pres. isomorphism.

$$\begin{tikzcd} X_1 \arrow[r, "T_1"] \arrow[d, "\Phi"'] & X_1 \arrow[d, "\phi"] \\ X_2 \arrow[r, "T_2"'] & X_2 \end{tikzcd}$$

More precisely $\exists A_1 \subset X_1$ full measure T_1-inv
$\exists A_2 \subset X_2$ full measure T_2-inv
s.t.
$$\begin{tikzcd} A_1 \arrow[r, "T_1"] \arrow[d, "\phi"'] & A_1 \arrow[d] \\ A_2 \arrow[r, "T_2"'] & A_2 \end{tikzcd}$$

Pick a finite $\overset{\text{meas}}{\text{partition}}$ α_2 of X_2.
$\alpha = \{A_1, \dots, A_r\}$.

Then we may find $B_1, \dots, B_r$ a meas partition of X_1 with
$m_1(B_i \triangle \Phi^{-1} A_i) = 0$, $\beta = \{B_1, \dots, B_r\}$

Given $i_0, \dots, i_{n-1} \in \{1, \dots, r\}$ we have
$$\Phi^{-1} \bigcap_{j=0}^{n-1} T_2^{-j} A_{i_j} = \bigcap_{j=0}^{n-1} T_1^{-j} B_{i_j}$$
So they have the same measure.
Thus
$$H\Big(\bigvee_{j=0}^{n-1} T_2^{-j} \alpha\Big) = H\Big(\bigvee T_1^{-j} \beta\Big)$$
and taking limits
$$h(T_2, \alpha) = h(T_1, \beta) \le h(\mathfrak{X}_1)$$
whence
$$h(\mathfrak{X}_2) = \sup_\alpha h(T_2, \alpha) \le h(\mathfrak{X}_1)$$

(b) $\mathfrak{X}_1$ and $\mathfrak{X}_2$ are factors of each other.

Theorem 6.19

(1) $h(T,\mathcal{A}) \le h(T,\mathcal{A}\vee\mathcal{E}) \le h(T,\mathcal{A}) + h(T,\mathcal{E})$

(2) $\boxed{h(T,\mathcal{A}) \le h(T,\mathcal{E}) + H(\mathcal{A}|\mathcal{E})}$

(3) $h(T,T^{-1}\mathcal{A}) = h(T,T^{-1}\mathcal{A})$ $\left(= h(T,T\mathcal{A})\ \text{if invertible}\right)$

(4) $\boxed{h(T,\mathcal{A}) = h(T,\mathcal{A}_p) \quad \forall p \ge 1}$

(5) When T is invertible:
$h(T,\mathcal{A}) = h(T, \bigvee_{i=-k}^{k} T^i\mathcal{A}) \quad \forall k \ge 0$

Prop 6.20
If $\mathcal{A}$ and $\mathcal{E}$ are finite sub-algebras of $\mathcal{B}_X$ then
$|h(T,\mathcal{A}) - h(T,\mathcal{E})| \le$
$d(\mathcal{A},\mathcal{E}) = H(\mathcal{A}|\mathcal{E}) + H(\mathcal{E}|\mathcal{A})$

In particular $h(T,\cdot)$ is a continuous function on the set of finite sub-algebras (mod 0)

proof Thm 6.19 (2).

Proof:

(1) Use $\bigvee_{i=0}^{n-1}(\mathcal{A}\vee\mathcal{E}) = (\bigvee_{i=0}^{n-1}\mathcal{A}) \vee (\bigvee_{i=0}^{n-1}\mathcal{E})$ to get

$$H(\mathcal{A}_n) \le H((\mathcal{A}\vee\mathcal{E})_n) = H(\mathcal{A}_n\vee\mathcal{E}_n) \le H(\mathcal{A}_n) + H(\mathcal{E}_n)$$

(2) $H(\mathcal{A}_n) \le H(\mathcal{A}_n \vee \mathcal{E}_n)$
$= H(\mathcal{E}_n) + H(\mathcal{A}_n|\mathcal{E}_n)$

$$H(\mathcal{A}_n|\mathcal{E}_n) \le \sum_{i=0}^{n-1} H(T^{-i}\mathcal{A}\,|\,\mathcal{E}_n) \le \sum_{i=0}^{n-1} H(T^{-i}\mathcal{A}\,|\,T^{-i}\mathcal{E}_n) = n\,H(\mathcal{A}|\mathcal{E})$$

(3) $H(\mathcal{A}_n) = H(T^{-1}\mathcal{A}_n) = H((T^{-1}\mathcal{A})_n)$

(4) $\mathcal{A}_p \vee \dots \vee T^{-n+1}\mathcal{A}_p = \mathcal{A}\vee T^{-1}\mathcal{A}\vee\dots\vee T^{-n-p+2}\mathcal{A}$
$= \mathcal{A}_{n+p-1}$

$$\frac{1}{n}H((\mathcal{A}_p)_n) = \frac{1}{n}H(\mathcal{A}_{n+p-1}) = \frac{n+p-1}{n}\,\frac{1}{n+p-1}H(\mathcal{A}_{n+p-1}) \to 1 \times H(T,\mathcal{A})$$
as $n\to\infty$

(5) $\bigvee_{i=-k}^{k} T^i\mathcal{A} = T^k\mathcal{A}_{2k+1}$

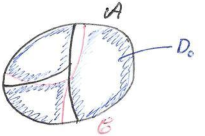
Lemma 6.21

Given a $k \in \mathbb{N}$, $\epsilon > 0$ there is $\delta = \delta(k, \epsilon) > 0$ such that:

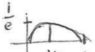
If $\xi_{\mathcal{A}} = (A_1, \dots, A_k)$ and $\xi_{\mathcal{C}} = (C_1, \dots, C_k)$ are two partitions of card k for which $\mu(A_i \triangle C_i) < \delta \ \forall i$

Then

$$d(\mathcal{A}, \mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{C}|\mathcal{A}) < \varepsilon.$$



proof: Recall that $\lim_{t \to 0^+} \phi(t) = 0$
$\phi(t) = -t \log t$.
$\phi(1) = 0$



We construct an auxilary partition: $\mathcal{D}$ consisting of

$$D_0 = \bigcup_{i=1}^{k} A_i \cap C_i \text{ and } A_i \cap C_j, \ i \neq j$$

Then $1 = \sum_i^k \mu(A_i) = \sum(\mu(A_i \cap C_i) + \mu(A_i \setminus C_i))$

$\le \mu(D_0) + k \cdot \delta$. so $\mu(D_0) \ge 1 - k\delta$.

Also $\mu(A_i \cap C_j) \le \mu(A_i \cap C_i^c) < \delta$.
for $i \neq j$

So if $1 - k\delta > 1/e$:

$$H(\mathcal{D}) = \sum_{i \neq j} \phi(\mu(A_i \cap C_j)) + \phi(D_0)$$

$$\le k \cdot (k-1)\, \phi(\delta) + \phi(1 - k\delta)$$

which goes to zero as $\delta \to 0$
So choose δ so that $k\delta < 1/e$ and the RHS $< \varepsilon/2$. Then by construction $\boxed{\mathcal{A} \vee \mathcal{C}} = \boxed{\mathcal{A} \vee \mathcal{D}}$ and

$$H(\mathcal{A} \vee \mathcal{C}) = H(\mathcal{A}) + H(\mathcal{C}|\mathcal{A}) =$$
$$H(\mathcal{A} \vee \mathcal{D}) \le H(\mathcal{A}) + H(\mathcal{D}) \le H(\mathcal{A}) + \varepsilon/2$$

so $H(\mathcal{C}|\mathcal{A}) < \varepsilon/2$. Similarly $H(\mathcal{A}|\mathcal{C}) < \varepsilon/2$

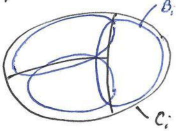
Prop 6.22

- $(X, \mathcal{B}_X, \mu)$ probability space.
- $\mathcal{B}_0 \subset \mathcal{B}_X$ a sub-algebra generating $\mathcal{B}_X = \sigma(\mathcal{B}_0)$

Then given any finite sub-alg $\mathcal{C} \subset \mathcal{B}_X$ and $\varepsilon > 0$ we may find a finite sub-alg $\mathcal{A} \subset \mathcal{B}_0$ for which

$$d(\mathcal{C}, \mathcal{A}) < \varepsilon.$$

proof



Let $\xi_{\mathcal{C}} = \{C_1, \dots, C_k\}$ be the partition of $\mathcal{C}$.

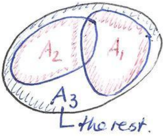
Let $\beta > 0$ (to be determined later) and pick $\forall i$

$B_i \in \mathcal{B}_0$ such that $\mu(B_i \Delta C_i) \leq \beta$.

The collection $\{B_1, \dots, B_k\}$ is an approximation of $\xi_{\mathcal{C}}$ but not a partition. Define

$$A_1 = B_1$$
$$A_2 = B_2 \cap B_1^c$$
$$A_3 = B_3 \cap B_2^c \cap B_1^c$$
$$\vdots$$
$$A_{k-1} = B_{k-1} \cap B_{k-2}^c \cap \dots \cap B_1^c$$

finally $A_k = B_{k-1}^c \cap B_{k-2}^c \cap \dots \cap B_1^c$



For notational simplicity write $|\mathcal{U}| = \mu(\mathcal{U})$.

$$|A_1 \Delta C_1| = |B_1 \Delta C_1| \leq \beta$$

For $j < k$ we have

$$|A_j \cap C_j^c| \leq |B_j \cap C_j^c| \leq \beta$$

$$C_j \cap A_j^c = C_j \cap (B_j^c \cup \bigcup_{i<j} B_i)$$
$$= C_j \cap (B_j^c \cup \bigcup_{i<j}(B_i \cap C_i^c))$$
$$\subset (C_j \cap B_j^c) \cup \bigcup_{i<j}(B_i \cap C_i^c)$$

So $|C_j \Delta A_j| \leq j \cdot \beta$

For $j = k$ we have

$$C_k \cap A_k^c = C_k \cap \bigcup_{i<k} B_i =$$
$$C_k \cap \bigcup_{i<k}(B_i \cap C_i^c) \quad \text{so}$$
$$|C_k \cap A_k^c| \leq \sum_{i<k} |B_i \cap C_i^c|$$

$$A_k \cap C_k^c = \left(\bigcap_{i<k} B_i^c\right) \cap \left(\bigcup_{j<k} C_j\right)$$
$$\subset \bigcup_{j<k}(C_j \cap B_j^c) \quad \text{and}$$
$$|A_k \cap C_k^c| \leq \sum_{j<k} |C_j \cap B_j^c|$$

which shows:

$$|C_k \Delta A_k| \leq (k-1)\beta$$

Choosing

$$\beta = \frac{1}{k-1}\delta(k, \epsilon)$$

with $\delta(k, \epsilon)$ as in Lemma 6.21 we obtain the wanted bound. //

Thm 6.23 [Kolonogorov-Sinai] 58-59

Let $(X, \mathcal{B}_X, \mu, T)$ be a meas pres transf of a probability space.

1. Suppose that $\mathcal{A} \subset \mathcal{B}_X$ is a finite subalgebra with
$$\bigvee_{i=0}^{\infty} T^{i}\mathcal{A} \doteq \mathcal{B}_X$$
Then $h_\mu(T) = h_\mu(T, \mathcal{A})$

2. Let T be invertible and suppose $\mathcal{A} \subset \mathcal{B}_X$ is finite and
$$\bigvee_{i=-\infty}^{\infty} T^{i}\mathcal{A} \doteq \mathcal{B}_X$$
Then $h_\mu(T) = h_\mu(T, \mathcal{A})$

1\.
proof: We let $\mathcal{A}_n = \bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}$
The hypothesis is ~~that~~
$\mathcal{B}_0 = \bigcup_{n \geq 1} \mathcal{A}_n$ generates
$\mathcal{B}_X = \sigma(\mathcal{B}_0)$.

Let $\epsilon > 0$ and $\xi_{\mathcal{C}} = \{C_1, \dots, C_k\}$ be a finite partition. We may then by Thm 6.22 find
~~A~~ $A_1, \dots, A_k \in \mathcal{B}_0$
such that $\eta = \{A_1, \dots, A_k\}$ satisfies
$$H(\mathcal{C} | \eta) < \varepsilon$$
Since k is finite we have
$\eta \subset \mathcal{A}_p$ for some $p \geq 1$

Then
$$\begin{aligned} h(T, \mathcal{C}) &\leq h(T, \mathcal{A}_p) + H(\mathcal{C}|\mathcal{A}_p) \\ &= h(T, \mathcal{A}) + H(\mathcal{C}|\mathcal{A}_p) \\ &\leq h(T, \mathcal{A}) + H(\mathcal{C}|\eta) \\ &\leq h(T, \mathcal{A}) + \varepsilon \end{aligned}$$
so as $\epsilon > 0$ was arbitrary
$$h_\mu(T) = \sup_{\mathcal{C} \text{ finite}} h_\mu(T, \mathcal{C}) \leq h_\mu(T, \mathcal{A}) \leq h_\mu(T).$$

2\. In this case we let $\mathcal{A}_n = \bigvee_{i=-n}^{n} T^{i}\mathcal{A}$ and note that again
$$\mathcal{B}_0 = \bigcup_{n \geq 0} \mathcal{A}_n$$
generates $\mathcal{B}_X$.
The remaining part is the same.

Entropy of a shift

$\Sigma_d^+ = \{ \underline{s} = (s_i)_{i\geq 0} \,;\, 1 \leq s_i \leq d \}$

$T = \sigma^+ : \Sigma_d^+ \to \Sigma_d^+$ shift

$\vec{p} = (p_1, \dots, p_d)$ probability vector

Define the measure on cylinders:

$\mu_{\vec{p}}([b_0 \dots b_{k-1}]) = p_{b_0} \cdots p_{b_{k-1}}$

$\{ \underline{s} \in \Sigma_d^+ : s_0 = b_0, \dots, s_{k-1} = b_{k-1} \}$

and let $\mu_{\vec{p}}$ be the Kolmogorov extended measure on Σ_d^+, invariant and ergodic under σ^+.

Let $\mathcal{A} = \{[1], \dots, [d]\}$ be the initial partition into elementary cylinders.

Then

$\mathcal{A}_k = \mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(k-1)}\mathcal{A} = \{[b_0 \dots b_{k-1}]\}$

is the collection of k-cylinders in Σ_d^+. They form a base for the topology, whence generate the Borel σ-algebra.

$\mathcal{B} = \sigma(\bigcup_{k\geq 1} \mathcal{A}_k)$

Since $\mu([b_0 \dots b_{k-1}]) = \mu([b_0]) \cdots \mu([b_{k-1}])$ we see that $= p_{b_0} \cdots p_{b_{k-1}}$, $\mathcal{A}, T^{-1}\mathcal{A}, \dots, T^{-(k-1)}\mathcal{A}_{k-1}$ are independent sub-algebras. Thus,

$H(\mathcal{A} \vee \dots \vee T^{-(k-1)}\mathcal{A}) = H(\mathcal{A}) + \dots + H(T^{-(k-1)}\mathcal{A})$

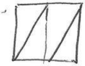
$= k\, H(\mathcal{A})$

$= k \sum_{i=1}^{d} p_i \log \frac{1}{p_i}$ and

K-S thm $\Rightarrow$ $h_{\mu_{\vec{p}}}(T) = \sum_1^d p_i \log \frac{1}{p_i}$

$T_2 x = 2x \mod 1$
Leb meas: λ

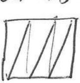
$(S^1, \lambda, T_2) \underset{\text{isom}}{\simeq} (\Sigma_2^+, \mu_{(\frac{1}{2},\frac{1}{2})}, \sigma^+)$



$h_\lambda(T_2) = \frac{1}{2}\log 2 + \frac{1}{2}\log 2$
$= \log 2$

$T_3 x = 3x \mod 1$, Leb λ

$(S^1, \lambda, T_3) \simeq (\Sigma_3^+, \mu_{(\frac{1}{3},\frac{1}{3},\frac{1}{3})}, \sigma^+)$



$h_\lambda(T_3) = \log 3$

Conclusion

(S^1, λ, T_2) and (S^1, λ, T_3) cannot be isomorphic.
