§10. Measure-theoretic entropy

We now study the measure-theoretic (a.k.a. Kolmogorov-Sinai) entropy of an invariant measure for a continuous map.

§10.1. Entropy of partitions

Assume X is a metric space and μ is a probability measure on X.

A <u>partition</u> of X is a collection of Borel sets $\xi = (A_j)_{j\in J}$, J at most countable, such that $\mu(A_j \cap A_k) = 0$ for $j \neq k$ and $\mu(X \setminus \bigcup_j A_j) = 0$.

Two partitions $(A_j)_{j\in J}$, $(B_j)_{j\in J}$ are identified if $\mu(A_j \Delta B_j) = 0$ $\forall j$. ($A \Delta B = (A\setminus B) \cup (B \setminus A)$)

We can think of a partition $\xi = (A_j)_{j\in J}$ a <u>function</u> $F_\xi : X \to J$ (defined μ-almost everywhere)

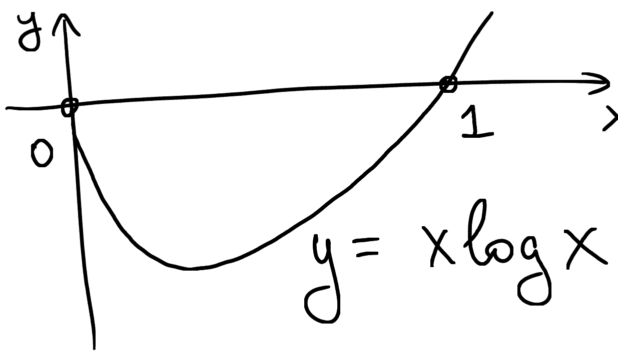
$F_\xi(x) =$ the unique $j \in J$ such that $x \in A_j$

discrete random variable...

Defn The *entropy* of a partition $\xi = (A_j)_{j \in J}$

$$H_\mu(\xi) := - \sum_{j \in J} \mu(A_j) \log \mu(A_j).$$

Here we define $0 \cdot \log 0 = 0$:



Note $x \log x$ is *convex*:

$$(x \log x)'' = (1 + \log x)' = \frac{1}{x} > 0$$

Another way to think about this is via the *information function*:

$$I_\xi : X \to \mathbb{R}, \quad I_\xi(x) = -\log \mu(A_{F_\xi(x)})$$

(here $A_{F_\xi(x)}$ is the element of the partition ξ containing x)

Then $$H_\mu(\xi) = \int_X I_\xi \, d\mu.$$

Henceforth we denote $A_\xi(x) := A_{F_\xi(x)}$.

Before going on, we ask

<u>Question:</u> Assume that $X = \{1, \dots, N\}$ and $\xi = \{\{1\}, \{2\}, \dots, \{N\}\}$.
Which μ <u>minimize</u> the entropy $H_\mu(\xi)$ and which μ <u>maximize</u> it?

<u>Answer:</u> Let $\mu(\{j\}) = c_j$, $c_j \geq 0$

$\sum_{j=1}^{N} c_j = 1$. Then

$$H_\mu(\xi) = -\sum_{j=1}^{N} c_j \log c_j.$$

<u>Minimize:</u> one of $c_j = 1$,
the rest $= 0$, $H_\mu(\xi) = 0$

<u>Maximize:</u> $c_j = \frac{1}{N}$ $\forall j$, $H_\mu(\xi) = \log N$.

Why is this the maximum?
Enough to check that at a maximal point of $H_\mu(\xi)$, we have $c_j = c_k$ $\forall j, k$.
For that it suffices to check that

↳

the function

$$(c_1, c_2) \mapsto -c_1 \log c_1 - c_2 \log c_2$$

on $\{c_1, c_2 \geq 0,\ c_1 + c_2 = \alpha\}$, $0 < \alpha \leq 1$ fixed

has only one maximal value, $c_1 = c_2 = \frac{\alpha}{2}$.

Put $c_1 = \alpha s$, $c_2 = \alpha(1-s)$, $0 \leq s \leq 1$,

then our function is

$$-\alpha s \log(\alpha s) - \alpha(1-s) \log[\alpha(1-s)]$$
$$= -\alpha \log \alpha - \alpha(s \log s + (1-s)\log(1-s))$$

and the function $s \mapsto -s \log s - (1-s)\log(1-s)$

has a unique maximum point on $[0,1]$, given by $s = \frac{1}{2}$.

<u>Conditional entropy</u>

Assume now that we are given two partitions ξ, η.

We want to define the <u>conditional entropy</u>

$H_\mu(\xi | \eta)$: the entropy of ξ assuming η is "known".

18.118
10-5

We use the <u>conditional measure</u>:

$$\mu(A|B) := \frac{\mu(A\cap B)}{\mu(B)}.$$

Define the <u>conditional entropy</u> as

$$H_\mu(\xi|\eta) = -\sum_{B\in\eta}\mu(B)\sum_{A\in\xi}\mu(A|B)\log\mu(A|B)$$

$$= -\sum_{\substack{A\in\xi\\ B\in\eta}}\mu(A\cap B)\log\mu(A|B)$$

$$= \int_X I_{\xi|\eta}\,d\mu \quad \text{where}$$

$$I_{\xi|\eta}(x) = -\log\mu(A|B) \quad \text{if } x\in A\cap B,\ A\in\xi,\ B\in\eta.$$

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

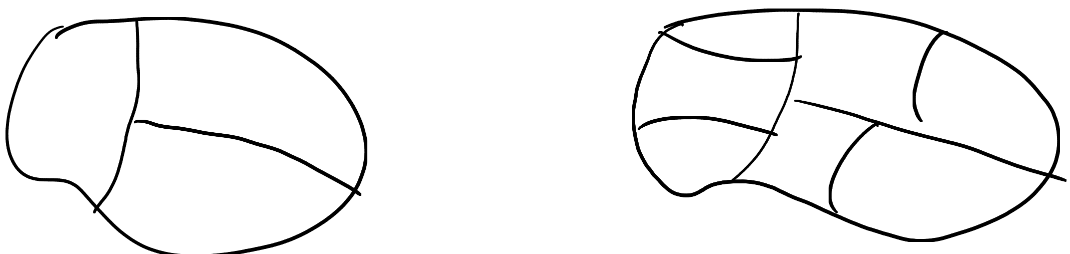
how much information we get from learing $F_\xi(x)$ if we already know $F_\eta(x)$

We say ξ,η are independent if

$$\mu(A\cap B) = \mu(A)\mu(B) \quad \forall A\in\xi,\ B\in\eta$$

For 2 partitions ξ, η
we write $\xi \leq \eta$
(or say η is a <u>refinement</u> of ξ)
if $\forall B \in \eta \quad \exists A \in \xi : \mu(B \setminus A) = 0$
(i.e. $B \subset A$ modulo a set of $\mu = 0$)

<u>Picture</u>:



ξ $\qquad\qquad$ η

Another way to think about it is:
F_ξ is a function of F_η.

<u>Lemma</u> We have $\forall$ partitions ξ, η

$0 \leq H_\mu(\xi | \eta) \leq H_\mu(\xi)$ and

$H_\mu(\xi|\eta) = H_\mu(\xi) \iff \xi, \eta$ are independent

$H_\mu(\xi|\eta) = 0 \iff \xi \leq \eta$

Proof

① $H_\mu(\xi|\eta) \geq 0$: we have

$$H_\mu(\xi,\eta) = -\sum_{\substack{A\in\xi \\ B\in\eta}} \mu(A\cap B) \log \mu(A|B).$$

Each term is ≥ 0, so $H_\mu(\xi|\eta) \geq 0$.

Next, assume that $H_\mu(\xi|\eta) = 0$.

Then $\forall A \in \xi$, $B \in \eta$, either

$$\mu(A\cap B) = 0 \quad \text{or} \quad \mu(A|B) = 1$$
$$(\text{i.e. } \mu(B\setminus A) = 0)$$

Thus $\xi \leq \eta$.

② $H_\mu(\xi|\eta) \leq H_\mu(\xi)$: we write, with $\Phi(x) := x \log x$, and Φ <u>strictly convex</u>,

$$H_\mu(\xi|\eta) = -\sum_{A\in\xi, B\in\eta} \mu(B)\, \Phi(\mu(A|B))$$
$$\leq -\sum_{A\in\xi} \Phi\Big(\sum_{B\in\eta} \mu(B)\mu(A|B)\Big)$$
$$= -\sum_{A\in\xi} \Phi(\mu(A)) = H_\mu(\xi). \qquad \square$$

Equality only if $\mu(A|B)$ is independent of B

i.e. $\mu(A|B) = \mu(A)$ i.e. $\mu(A\cap B) = \mu(A)\mu(B)$

Note: if $\eta = \{X\}$ is the trivial partition then

$$H_\mu(\xi|\eta) = H_\mu(\xi).$$

Similarly to the lemma above we can see that for any 3 partitions ξ, η, ζ,

$$\eta \leq \zeta \Rightarrow H(\xi|\zeta) \leq H(\xi|\eta)$$

(the lemma was for the case when η is trivial)

Roughly speaking, ζ gives us more knowledge so ξ can only add less information to ζ than to η.

Joint partition:

if ξ, η are partitions then define the partition $\xi \vee \eta$ by

$$\xi \vee \eta = \{A \cap B \mid A \in \xi, B \in \eta\}.$$

This corresponds to the random variable

$$F_{\xi \vee \eta}(x) = (F_\xi(x), F_\eta(x)).$$

Lemma We have $\forall$ partitions ξ, η,

$$H_\mu(\xi \vee \eta) = H_\mu(\xi | \eta) + H_\mu(\eta). \qquad (1)$$

Moreover,

$$H_\mu(\xi \vee \eta) \leq H_\mu(\xi) + H_\mu(\eta). \qquad (2)$$

Proof (1): we write

$$H_\mu(\xi \vee \eta) = \sum_{A \in \xi, B \in \eta} \mu(A \cap B) \log \mu(A \cap B)$$

$$= \sum_{B \in \eta} \mu(B) \sum_{A \in \xi} \mu(A|B) \log \left[\mu(B) \mu(A|B)\right]$$

$$= \sum_{B \in \eta} \mu(B) \left[\log \mu(B) + \sum_{A \in \xi} \mu(A|B) \log \mu(A|B)\right]$$

$$= H_\mu(\eta) + H_\mu(\xi | \eta).$$

(2): this follows from (1) and the previously shown inequality

$$H_\mu(\xi | \eta) \leq H_\mu(\xi).$$

§10.2. Entropy of a map

Let X be a metric space,
$\varphi : X \circlearrowleft$ be a (Borel measurable) map,
and μ be a φ-invariant probability measure on X.

Fix some partition ξ on X.
Define the refined partition

$$\xi^{(n)} = \xi^{(n)}_{\varphi} := \xi \vee \varphi^{-1}(\xi) \vee \dots \vee \varphi^{-(n-1)}(\xi).$$

Here $\varphi^{-j}(\xi)$ is the partition

$$\varphi^{-j}(\xi) := \{ \varphi^{-j}(A) : A \in \xi \}$$

Note: the random variable corresponding to $\xi^{(n)}$ is

$$F_{\xi^{(n)}}(x) = \left(F_{\xi}(x), F_{\xi}(\varphi(x)), \dots, F_{\xi}(\varphi^{n-1}(x)) \right)$$

i.e. it encodes which elements of the partition ξ have the points $x, \varphi(x), \dots, \varphi^{n-1}(x)$.

Consider the entropies $H_\mu(\xi^{(n)})$. They are <u>subadditive</u>:

$\forall n, m \geq 0$,

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

Indeed, we have

$$\xi^{(n+m)} = \xi^{(n)} \vee \varphi^{-n}(\xi^{(m)}),$$

so $H_\mu(\xi^{(n+m)}) \leq H_\mu(\xi^{(n)}) + H_\mu(\varphi^{-n}(\xi^{(m)}))$

and $H_\mu(\varphi^{-n}(\xi^{(m)})) = H_\mu(\xi^{(m)})$:

if φ is measure preserving and ξ is any partition, then $H_\mu(\varphi^{-1}(\xi)) = H_\mu(\xi)$.

Subadditivity gives existence of the limit

$$h_\mu(\varphi, \xi) := \lim_{n \to \infty} \frac{1}{n} H_\mu(\xi_\varphi^{(n)})$$

<u>measure theoretic entropy</u> of φ relative to the partition ξ.

Definition The measure-theoretic entropy of φ w.r.t. μ is

$$h_\mu(\varphi) := \sup\{h_\mu(\varphi, \xi) \mid \xi \text{ a finite partition}\}.$$

Examples of computation of $h_\mu(\varphi, \xi)$:

① $X = S^1$, $\varphi(x) = x + r \mod \mathbb{Z}$ ($r \in \mathbb{R}$ fixed),

$\mu =$ Lebesgue, $\xi = \{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}$.

The partition $\xi^{(n)}$ consists of various intersections of the $2n$ intervals $\varphi^{-j}([0, \frac{1}{2}])$ and $\varphi^{-j}([\frac{1}{2}, 1])$, $0 \le j < n$.

Then $\xi^{(n)}$ has $\le 2n$ elements, which gives (looking at the question in §10.1) that

$$H_\mu(\xi^{(n)}) \le \log(2n)$$

Thus $h_\mu(\varphi, \xi) = 0$.

② $X = S^1 = \mathbb{R}/\mathbb{Z}$, $\varphi(x) = 2x \mod \mathbb{Z}$.

$\mu =$ Lebesgue, $\xi = \{[0,\frac{1}{2}], [\frac{1}{2},1]\}$.

Call $\xi = \{A_0, A_1\}$, where $A_0 = [0,\frac{1}{2}]$
$A_1 = [\frac{1}{2},1]$.

Then $\xi \vee \varphi^{-1}(\xi) \vee \cdots \vee \varphi^{-(n-1)}(\xi) = \xi^{(n)}$

consists of sets $A_{\vec{w}}$, $\vec{w} \in \{0,1\}^n$:

$$x \in A_{w_0 w_1 \cdots w_{n-1}} \iff \begin{cases} x \in A_{w_0} \\ \varphi(x) \in A_{w_1} \\ \vdots \\ \varphi^{n-1}(x) \in A_{w_{n-1}} \end{cases}$$

$(w_e \in \{0,1\})$

Up to a measure 0 set, $A_{w_0 \cdots w_{n-1}}$ is just the set of $x \in [0,1]$ whose binary expansion starts with $0.w_0 w_1 \cdots w_{n-1}$, that is $A_{w_0 \cdots w_{n-1}}$ is an interval of length 2^{-n}.

So $\forall A \in \xi^{(n)}$, $\mu(A) = 2^{-n}$, which gives $H_\mu(\xi^{(n)}) = n \log 2$

and thus $h_\mu(\varphi, \xi) = \log 2$.

③ <u>Hyperbolic toral automorphisms</u>:

$X = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, $\varphi(x) = Ax \bmod \mathbb{Z}^2$,

$A \in SL(2, \mathbb{Z})$ hyperbolic.

Take μ = Lebesgue measure.

Let λ, λ^{-1} be the eigenvalues of A, with $|\lambda| > 1$.

We show that $\exists\, \varepsilon_0 > 0$: if ξ is a partition with each element having diameter $< \varepsilon_0$, then

$$\boxed{h_\mu(\varphi, \xi) \geq \log|\lambda|.}$$

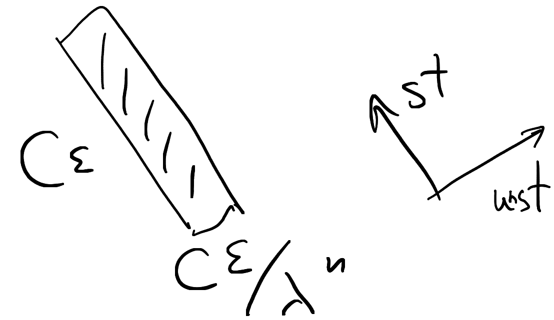
(So then, also $h_\mu(\varphi) \geq \log|\lambda|$)

(Later we might prove that $h_\mu(\varphi) = \log|\lambda|$...)

To see this, note that for any 2 points x, y lying in the same element of the refined partition $\xi^{(n)}$, we have $d(\varphi^j(x), \varphi^j(y)) \leq \varepsilon_0$ for $j = 0, 1, \dots, n-1$.

So (see §9) if ε_0 is small enough, then the <u>unstable</u> distance from x to y is $\leq \dfrac{C\varepsilon_0}{\lambda^n}$.

That is, each element $A \in \xi^{(n)}$ is contained in an unstable rectangle:



Thus $\mu(A) \le \dfrac{C^2\varepsilon^2}{|\lambda|^n}$.

Recalling the formula for the entropy $H_\mu(\xi^{(n)})$, we see that

$$H_\mu(\xi^{(n)}) \ge -\log \frac{C^2\varepsilon^2}{|\lambda|^n} = n\log|\lambda| + O(1) \quad \text{as } n\to\infty$$

So $h_\mu(\varphi,\xi) \ge \log|\lambda|$.

---

§10.3. Generating partitions

It turns out that $h_\mu(\varphi) = h_\mu(\varphi,\xi)$ for a sufficiently fine partition ξ in many cases.

Denote by $\mathcal{P}_m$ the set of all partitions of X into m sets.

We define the following <u>metric</u> on P_m:

$$d(\xi,\eta) := \min_{\sigma} \sum_{A\in\xi} \mu(A \,\Delta\, \sigma(A))$$

where σ goes over all bijections $\sigma: \xi \to \eta$.

<u>Defn</u> If $\varphi: X \circlearrowleft$ is a map preserving a probability measure μ and ξ is a partition, we say that ξ is:

- a <u>one-sided generator</u> for φ, if $\forall$ finite partition η, $\forall \delta > 0$ $\exists\, n$ and a partition $\zeta \leq \xi^{(n)}$ (here $\xi^{(n)} = \xi \vee \varphi^{-1}(\xi) \vee \cdots \vee \varphi^{-(n-1)}(\xi)$) such that $\boxed{d(\eta,\zeta) \leq \delta.}$

That is, any finite partition is well-approximated by partitions subordinate to $\xi^{(n)}$.

18.118
10-17

- if φ is invertible, we call ξ a <u>generator</u> for φ, if the same property holds with $\xi^{(n)}$ replaced by

$$\bigvee_{j=-n}^{n} \varphi^j(\xi).$$

We will show

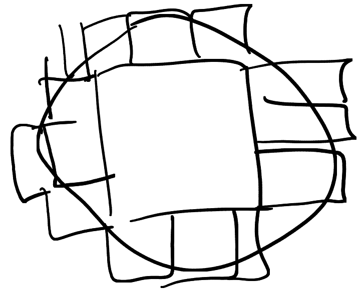
<u>Thm</u> Assume that ξ is a one-sided generator, or φ is invertible & ξ is a generator. Then $\boxed{h_\mu(\varphi) = h_\mu(\varphi, \xi)}.$

<u>Example:</u> Assume that μ is nonatomic and $\max_{A \in \xi^{(n)}} \operatorname{diam}(A) \longrightarrow 0$ as $n \to \infty$.

Then ξ is a one-sided generator.

(Similarly for φ invertible & ξ a generator)

Will skip the proof but this works similarly to approximating arbitrary Lebesgue measurable sets by unions of small cubes:



So e.g.

- for irrational shift on S^1 (example ①), $\xi = [0, \frac{1}{2}], [\frac{1}{2}, 1]$ is a 1-sided generator (so $h_\mu(\varphi) = 0$)
- for the map $x \mapsto 2x$ on S^1, the same ξ is a 1-sided generator (so $h_\mu(\varphi) = \log 2$)
- for the cat map, any ξ consisting of sets with small diameter is a generator (but it is not a 1-sided generator)

We now start the proof of Thm, with

<u>Lemma</u> We have $\forall$ partitions ξ and η

$$h_\mu(\varphi, \eta) \leq h_\mu(\varphi, \xi) + H_\mu(\eta \mid \xi).$$

<u>Proof</u> 1 Denote

$$\xi^{(n)} := \xi \vee \varphi^{-1}(\xi) \vee \dots \vee \varphi^{-(n-1)}(\xi)$$

$$\eta^{(n)} := \eta \vee \varphi^{-1}(\eta) \vee \dots \vee \varphi^{-(n-1)}(\eta).$$

Then $H_\mu(\xi^{(n)}) \leq H_\mu(\xi^{(n)} \vee \eta^{(n)}) =$

$$= H_\mu(\eta^{(n)}) + H_\mu(\xi^{(n)} \mid \eta^{(n)}).$$

Recalling the definition of h_μ, we see that it suffices to prove the inequality

$$(*) \quad H_\mu(\xi^{(n)} \mid \eta^{(n)}) \leq n H_\mu(\xi \mid \eta).$$

2 We show $(*)$ by induction on n.

$n = 1$ is immediate.

For $n \geq 2$, we use the identity (valid for any 3 partitions ξ, η, ζ)

$$(**)\quad H_\mu(\xi \vee \eta \mid \zeta) = H_\mu(\xi \mid \zeta) + H_\mu(\eta \mid \xi \vee \zeta).$$

To check $(**)$ we compute

$$H_\mu(\xi \vee \eta \mid \zeta) = \sum_{A\in\xi, B\in\eta, C\in\zeta} \mu(A\cap B\cap C) \log \frac{\mu(A\cap B\cap C)}{\mu(C)}$$

$$= \sum_{A,B,C} \mu(A\cap B\cap C) \log \left[\frac{\mu(A\cap B\cap C)}{\mu(A\cap C)} \cdot \frac{\mu(A\cap C)}{\mu(C)} \right]$$

$$= \underbrace{\sum_{A,B,C} \mu(A\cap B\cap C) \log \frac{\mu(A\cap B\cap C)}{\mu(A\cap C)}}_{H(\eta \mid \xi \vee \zeta)} +$$

$$+ \underbrace{\sum_{A,C} \mu(A\cap C) \log \frac{\mu(A\cap C)}{\mu(C)}}_{H_\mu(\xi \mid \zeta)}.$$

③ Coming back to the proof of $(*)$, we write (as $\xi^{(n)} = \xi \vee \varphi^{-1}(\xi^{(n-1)})$)

$$H_\mu(\xi^{(n)} \mid \eta^{(n)}) =$$

$$= H(\xi \mid \eta^{(n)}) + H(\varphi^{-1}(\xi^{(n-1)}) \mid \xi \vee \eta^{(n)})$$

$$\le H(\xi \mid \eta) + H(\varphi^{-1}(\xi^{(n-1)}) \mid \varphi^{-1}(\eta^{(n-1)}))$$

(as $\eta \le \eta^{(n)}$, $\varphi^{-1}(\eta^{(n-1)}) \le \xi \vee \eta^{(n)}$)

$$= H(\xi \mid \eta) + H(\xi^{(n-1)} \mid \eta^{(n-1)})$$ and we proceed by induction on n. $\square$

as φ is measure preserving

<u>Corollary</u>: if $\eta \leq \xi$ then

$$H_\mu(\eta|\xi) = 0$$ and thus

$$h_\mu(\varphi, \eta) \leq h_\mu(\varphi, \xi).$$

<u>Lemma</u> We have $\forall m \geq 0$

$$h_\mu(\varphi, \xi^{(m)}) = h_\mu(\varphi, \xi).$$

If φ is invertible, same is true with $\xi^{(m)}$ replaced by $\bigvee_{\ell=-m}^{m} \varphi^\ell(\xi)$.

<u>Proof</u> We just show the first statement: the second one holds since

$$\bigvee_{\ell=-m}^{m} \varphi^\ell(\xi) = \varphi^m(\xi^{(2m+1)}).$$

The n-th refinement of $\xi^{(m)}$ is

$$(\xi^{(m)})^{(n)} = \bigvee_{j=0}^{n-1} \varphi^{-j}(\xi^{(m)}) = \bigvee_{j=0}^{n-1} \bigvee_{\ell=0}^{m-1} \varphi^{-j}(\varphi^{-\ell}(\xi))$$

$$= \bigvee_{j=0}^{n+m-1} \varphi^{-j}(\xi) = \xi^{(m+n)}.$$ Then

$$h_\mu(\varphi, \xi^{(m)}) = \lim_{n\to\infty} \frac{H_\mu(\xi^{(m+n)})}{n} = \lim_{n\to\infty} \frac{H_\mu(\xi^{(m+n)})}{m+n}$$

$$= h_\mu(\varphi, \xi). \qquad \square$$

<u>Lemma</u> Fix $m \geq 1$ and let $\mathcal{P}_m$ be the set of partitions with m elements.

Then $\forall\, \varepsilon > 0 \;\; \exists\, \delta > 0$:

$\forall\, \xi, \eta \in \mathcal{P}_m$, if $\boxed{d(\xi,\eta) \leq \delta}$

then $\boxed{H_\mu(\eta|\xi) \leq \varepsilon.}$

(i.e. once m is fixed, if the sets in ξ, η are close to each other then $H(\eta|\xi)$ is small).

<u>Proof</u> We may write $\xi = (A_j)_{j=1}^m$, $\eta = (B_j)_{j=1}^m$

so that $\sum_{j=1}^m \mu(A_j \Delta B_j) \leq \delta$ (if $d(\xi,\eta) \leq \delta$).

Denote $\alpha_j := \dfrac{\mu(A_j \setminus B_j)}{\mu(A_j)} = \mu(B_j^c \,|\, A_j)$, then

$$\sum_{j=1}^m \mu(A_j)\alpha_j = \sum_{j=1}^m \mu(A_j \setminus B_j) \leq \sum_{j=1}^m \mu(A_j \Delta B_j) \leq \delta.$$

We have

$$\boxed{H_\mu(\eta|\xi) = -\sum_{j,k=1}^m \mu(A_j \cap B_k) \log \mu(B_k \,|\, A_j)}$$

and $\mu(B_j \,|\, A_j) = 1 - \alpha_j$

We split the sum into 2 parts.

$\underline{\underline{j=k}}$: get $-\sum_{j=1}^{m} \mu(A_j \cap B_j) \log \mu(B_j \mid A_j) =$

$$= -\sum_{j=1}^{m} \mu(A_j)(1-\alpha_j)\log(1-\alpha_j)$$

$\underline{\underline{j\neq k}}$: get $-\sum_{j=1}^{m} \mu(A_j) \sum_{k\neq j} \mu(B_k \mid A_j) \log \mu(B_k \mid A_j)$

$$= -\sum_{j=1}^{m} \mu(A_j)\alpha_j \sum_{k\neq j} \frac{\mu(B_k \mid A_j)}{\alpha_j} \log\left[\frac{\mu(B_k \mid A_j)}{\alpha_j}\cdot \alpha_j\right]$$

$$= -\sum_{j=1}^{m} \mu(A_j)\alpha_j \sum_{k\neq j} c_{jk} \log(\alpha_j \cdot c_{jk})$$

where $c_{jk} := \dfrac{\mu(B_k \mid A_j)}{\alpha_j} = \dfrac{\mu(B_k \cap A_j)}{\mu(A_j \setminus B_j)}$

and $\sum_{k\neq j} c_{jk} = 1$ (as $\bigsqcup_{k\neq j} (B_k \cap A_j) = A_j \setminus B_j$)

Thus $-\sum_{k\neq j} c_{jk} \log(\alpha_j \cdot c_{jk}) = -\log \alpha_j - \sum_{k\neq j} c_{jk} \log c_{jk}$

$\leq -\log \alpha_j + \log(m-1)$ (by the extreme case discussed in §10.1)

So the contribution of $j\neq k$ is

$$\leq -\sum_{j=1}^{m} \mu(A_j)\alpha_j\left(-\log \alpha_j + \log(m-1)\right).$$

Putting these together, we get $H_\mu(\eta|\xi) \leq$

$$\leq \sum_{j=1}^{m} \mu(A_j)\left(-(1-\alpha_j)\log(1-\alpha_j) - \alpha_j \log \alpha_j + \alpha_j \log(m-1)\right)$$

$$= \sum_{j=1}^{m} \mu(A_j) \cdot \beta_j \quad \text{where}$$

$$\beta_j := -(1-\alpha_j)\log(1-\alpha_j) - \alpha_j \log \alpha_j + \alpha_j \log(m-1)$$

Note that $\boxed{\beta_j \leq \log m,}$ as

$$\beta_j = (1-\alpha_j)\log\frac{1}{1-\alpha_j} + \alpha_j \log\frac{m-1}{\alpha_j} \leq$$ (as log is a concave function)

$$\leq \log\left(1-\alpha_j \cdot \frac{1}{1-\alpha_j} + \alpha_j \cdot \frac{m-1}{\alpha_j}\right) = \log m.$$

Put $\psi(x) := -(1-x)\log(1-x) - x\log x$, then

ψ is increasing on $[0, \frac{1}{2}]$.

We now split the sum into big & small $\mu(A_j)$:

$$H_\mu(\eta|\xi) \leq \sum_{\mu(A_j) > \sqrt{\delta}} \mu(A_j) \cdot \beta_j + \sum_{\mu(A_j) \leq \sqrt{\delta}} \mu(A_j) \cdot \beta_j$$

$$= \mathrm{I} + \mathrm{II}.$$

If $\mu(A_j) > \sqrt{\delta}$, then, since $\sum_j \mu(A_j)\alpha_j \leq \delta$, we have $\alpha_j \leq \sqrt{\delta} \Rightarrow \psi(\alpha_j) \leq \psi(\sqrt{\delta})$.

Thus $I = \sum_{\mu(A_j) > \sqrt{\delta}} \mu(A_j) \cdot (\psi(\alpha_j) + \alpha_j \log(n-1))$

$\leq \psi(\sqrt{\delta}) + \sqrt{\delta} \log(n-1)$

And $II \leq \sum_{\mu(A_j) \leq \sqrt{\delta}} \mu(A_j) \log n \leq \sqrt{\delta} \cdot n \log n.$

So $H_\mu(\eta | \xi) \leq \psi(\sqrt{\delta}) + \sqrt{\delta}(n \log n + \log(n-1))$

which goes to 0 as $\delta \to 0$. $\square$
(for n fixed)

We can now give

<u>Proof of Thm</u>

We assume ξ is a one-sided generator
(the case of φ invertible & ξ a generator is handled similarly).

Since $h_\mu(\varphi) := \sup\{h_\mu(\varphi, \eta) : \eta$ a finite partition$\}$,

to show that $h_\mu(\varphi) = h_\mu(\varphi, \xi)$
it suffices to prove that $\forall$ finite partition η we have

$$h_\mu(\varphi, \eta) \leq h_\mu(\varphi, \xi).$$

Put $n :=$ number of elements in η.

Take arbitrary $\varepsilon>0$. Let $\delta>0$ be from the last lemma.
Since ξ is a one-sided generator,
$\exists\, n\geq 0$ and a partition $\boxed{\zeta\leq\xi^{(n)}}$
such that $\boxed{d(\eta,\zeta)\leq\delta}$.
By the last lemma, we have
$\boxed{H_\mu(\eta|\zeta)\leq\varepsilon}$.

Now,

$$h_\mu(\varphi,\eta)\leq h_\mu(\varphi,\zeta)+H_\mu(\eta|\zeta)$$

(as $\zeta\leq\xi^{(n)}$) $\leq h_\mu(\varphi,\xi^{(n)})+\varepsilon$

$$\leq h_\mu(\varphi,\xi)+\varepsilon$$

This is true $\forall\,\varepsilon>0$, so $h_\mu(\varphi,\eta)\leq h_\mu(\varphi,\xi)$ as needed. $\square$