§ 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 $\mu$ is a probability measure on $X$.

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

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 $\mathcal{E} = \{A_j\}_{j \in J}$ a function $F_{\mathcal{E}} : X \to J$ (defined $\mu$-almost everywhere)

$$F_{\mathcal{E}}(x) = \text{the unique } j \in J \text{ such that } x \in A_j$$

discrete random variable...
**Defn.** The entropy of a partition $\mathcal{E} = (A_j)_{j \in J}$

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

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:

$${\mathcal I}_\mathcal{E} : X \to \mathbb{R}, \quad {\mathcal I}_\mathcal{E}(x) = -\log \mu(A_{F_{\mathcal{E}}(x)})$$

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

Then $H_\mu(\mathcal{E}) = \int_X {\mathcal I}_\mathcal{E} \, d\mu$.

Henceforth we denote $A_{\mathcal{E}}(x) := A_{F_{\mathcal{E}}(x)}$. 
Before going on, we ask

**Question:** Assume that \( X = \{1, \ldots, N\} \) and \( \xi = \{\xi_1, \xi_2, \ldots, \xi_N\} \).

Which \( \mu \) **minimize** the entropy \( H_\mu(\xi) \) and which \( \mu \) **maximize** it?

**Answer:** Let \( \mu(\{j\}) = c_j, \ c_j \geq 0 \)

\[
\sum_{j=1}^{n} c_j = 1. \text{ Then}
\]

\[
H_\mu(\xi) = -\sum_{j=1}^{n} c_j \log c_j.
\]

**Minimize:** one of \( c_j = 1 \), the rest = 0, \( H_\mu(\xi) = 0 \)

**Maximize:** \( c_j = \frac{1}{N} \ \forall j, \ H_\mu(\xi) = \log N \).

Why is this the maximum?

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

\( L \)
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\] 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 \(s = \frac{1}{2}\).

**Conditional entropy**

Assume now that we are given two partitions \(\xi, \eta\).

We want to define the **conditional entropy**

\(H_{\mu}(\xi|\eta)\): the entropy of \(\xi\)

assuming \(\eta\) is "known".
We use the conditional measure:

\[ m(AB) = \frac{\mu(AB)}{\mu(B)} \]

Define the conditional entropy as:

\[ H_m(\xi | \eta) = -\sum_{B \in \mathcal{B}} \mu(B) \sum_{A \in \mathcal{A}} \mu(AB) \log \mu(AB) \]

\[ = -\sum_{A \in \mathcal{A}} \mu(AB) \log \mu(AB) \]

\[ = \int_{\mathcal{X}} I_{\xi | \eta}(x) \, d\mu(x) \text{ where} \]

\[ I_{\xi | \eta}(x) = -\log \mu(AB) \text{ if } x \in A \cap B \]

\[ = -\log \frac{\mu(AB)}{\mu(B)} \text{ if } B \in \mathcal{B} \]

This is how much information we get from leaving \( F_\xi(x) \)

if we already know \( F_\eta(x) \)

We say \( \xi, \eta \) are independent if:

\[ \mu(AB) = \mu(A) \mu(B) \quad \forall A \in \mathcal{A}, B \in \mathcal{B} \]
For 2 partitions \( \xi, \eta \), we write \( \xi \leq \eta \) (or say \( \eta \) is a refinement of \( \xi \)) if \( \forall B \in \eta \exists A \in \xi : \mu(B \Delta A) = 0 \) (i.e. \( B \cap A \) modulo a set of \( \mu = 0 \))

Another way to think about it is: \( F_\xi \) is a function of \( F_\eta \).

**Lemma** We have 2 partitions \( \xi, \eta \)

\[ 0 \leq H_\mu(\xi | \eta) \leq H_\mu(\xi) \] and

\[ H_\mu(\xi | \eta) = H_\mu(\xi) \iff \xi, \eta \text{ are independent} \]

\[ H_\mu(\xi | \eta) = 0 \iff \xi \leq \eta \]
Proof

1. \( H_\mu(\xi|\eta) \geq 0 \): we have

\[
H_\mu(\xi,\eta) = - \sum_{\xi \in \Xi \atop \beta \in \eta} \mu(\xi \cap \beta) \log \mu(\xi | \beta).
\]

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

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

Then \( \forall \xi \in \Xi, \beta \in \eta \), either

\[
\mu(\xi \cap \beta) = 0 \text{ or } \mu(\xi | \beta) = 1
\]

(i.e. \( \mu(\beta | \xi) = 0 \)).

Thus \( \xi \leq \eta \).

2. \( H_\mu(\xi|\eta) \leq H_\mu(\xi) \): we write,

with \( \Phi(x) = x \log x \), and \( \Phi \) strictly convex,

\[
H_\mu(\xi|\eta) = - \sum_{\xi \in \Xi \atop \beta \in \eta} \mu(\xi) \Phi(\mu(\beta | \xi))
\]

\[
\leq - \sum_{\xi \in \Xi} \Phi\left( \sum_{\beta \in \eta} \mu(\beta) \mu(\beta | \xi) \right)
\]

\[
= - \sum_{\xi \in \Xi} \Phi\left( \mu(\xi) \right) = H_\mu(\xi).
\]

Equality only if \( \mu(\xi | \beta) \) is independent of \( \beta \)

i.e. \( \mu(\xi | \beta) = \mu(\xi) \) i.e. \( \mu(\xi \cap \beta) = \mu(\xi) \mu(\beta) \).
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 \( \xi, \eta, \zeta \),
\[
\eta \leq \zeta \implies H(\xi | \zeta) \leq H(\xi | \eta)
\]
(the lemma was for the case when \( \eta \) is trivial).

Roughly speaking, \( \xi \) gives us more knowledge so \( \xi \) can only add less information to \( \zeta \) than to \( \eta \).

Joint partition:
if \( \xi, \eta \) are partitions then define
the partition \( \xi \vee \eta \) by
\[
\xi \vee \eta = \{ AB | A \in \xi, \ B \in \eta \}
\]
This corresponds to the random variable
\[
F_{\xi \vee \eta} (x) = (F_\xi (x), F_\eta (x)).
\]
Lemma: We have, for partitions $\xi, \eta$,

$$H_\mu (\xi \upharpoonright \eta) = H_\mu (\xi | \eta) + H_\mu (\eta). \quad (1)$$

Moreover,

$$H_\mu (\xi \upharpoonright \eta) \leq H_\mu (\xi) + H_\mu (\eta). \quad (2)$$

Proof (1): we write

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

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

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

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

(2): this follows from (1)

and the previously shown inequality

$$H_\mu (\xi \upharpoonright \eta) \leq H_\mu (\xi).$$
10.2. Entropy of a map

Let \( X \) be a metric space, 
\( \varphi : X \to X \) be a (Borel measurable) map, 
and \( \mu \) be a \( \varphi \)-invariant probability measure on \( X \).

Fix some partition \( \mathcal{E} \) on \( X \). Define the refined partition 
\[
\mathcal{E}^{(n)} = \mathcal{E}^{(n)}_\varphi := \mathcal{E} \cup \varphi^{-1}(\mathcal{E}) \cup \ldots \cup \varphi^{-(n-1)}(\mathcal{E}).
\]

Here \( \varphi^{-1}(\mathcal{E}) \) is the partition 
\[
\varphi^{-1}(\mathcal{E}) := \{ \varphi^{-1}(A) : A \in \mathcal{E} \}.
\]

Note: The random variable corresponding to \( \mathcal{E}^{(n)} \) is 
\[
F_{\mathcal{E}^{(n)}}(x) = (F_{\mathcal{E}}(x), F_{\mathcal{E}}(\varphi(x)), \ldots, F_{\mathcal{E}}(\varphi^{n-1}(x)))
\]
i.e. it encodes which elements of the partition \( \mathcal{E} \) have the points 
\( x, \varphi(x), \ldots, \varphi^{n-1}(x) \).
Consider the entropies $H_\mu(\xi^{(n)})$. They are subadditive:

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

Indeed, we have

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

so

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

and

$$H_\mu(\phi^{-n}(\xi^{(m)})) = H_\mu(\xi^{(m)}).$$

if $\phi$ is measure preserving and $\xi$ is any partition, then $H_\mu(\phi^{-1}(\xi)) = H_\mu(\xi)$. Subadditivity gives existence of the limit

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

measure theoretic entropy of $\phi$ relative to the partition $\xi$. 

18.118
10-11
Definition. The measure-theoretic entropy of $\varphi$ w.r.t. $\mu$ is

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

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

1. $X = \mathbb{S}^1$, $\varphi(x) = x + r \text{ mod } 2$ ($r \in \mathbb{R}$ fixed), $\mu = \text{Lebesgue}$, $\xi = \{ [0, \frac{1}{2}], [\frac{1}{2}, 1] \}$. The partition $\xi^{(n)}$ consists of various intersections of the $2n$ intervals $\varphi^{-j}([0, \frac{1}{2}])$ and $\varphi^{-j}([\frac{1}{2}, 1])$, $0 \leq j \leq n$. Then $\xi^{(n)}$ has $\leq 2n$ elements, which gives (looking at the question in §10.1) that

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

Thus $h_{\mu}(\varphi, \xi) = 0$. 
\( X = \mathbb{S} = \mathbb{R}/\mathbb{Z}, \quad \varphi(x) = 2x \mod \mathbb{Z}. \)

\( \mu = \text{Lebesgue}, \quad \xi = \left\{ \left[ 0, \frac{1}{2} \right], \left[ \frac{1}{2}, 1 \right] \right\}. \)

Call \( \xi = \{ A_0, A_1 \}, \) where \( A_0 = \left[ 0, \frac{1}{2} \right], \quad A_1 = \left[ \frac{1}{2}, 1 \right] \).

Then \( \xi \supseteq \bigcup_{n=1}^{\infty} \varphi^{-n}(\xi) \) consists of sets \( A^1_n, \) \( \bar{w} \in \{0, 1\}^n: \)

\[ x \in A^1_{w_0 w_1 \ldots w_{n-1}} \iff \begin{cases} x \in A_{w_0} \\ \varphi(x) \in A_{w_1} \\ \vdots \\ \varphi^{n-1}(x) \in A_{w_{n-1}} \end{cases} \quad (w \in \{0, 1\}^n) \]

Up to a measure 0 set, \( A_{w_0 \ldots w_{n-1}} \) is just the set of \( x \in [0, 1] \) whose binary expansion starts with \( 0, w_0 w_1 \ldots w_{n-1}, \) that is \( A_{w_0 \ldots w_{n-1}} \) 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. Hyperbolic toral automorphisms:

\[ X = \Gamma^2 = \mathbb{R}^2 / \mathbb{Z}^2, \quad \varphi(x) = Ax \mod \mathbb{Z}^2, \]

\[ A \in \text{SL}(2, \mathbb{Z}) \text{ hyperbolic}. \]

Take \( \mu = \text{Lebesgue measure}. \)

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

We show that \( \exists \epsilon_0 > 0 \): if \( \xi \) is a partition with each element having diameter \( < \epsilon_0 \), then

\[ h_{\mu}(\varphi, \xi) \geq \log |\lambda|. \]

(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

\[ d(\varphi_j(x), \varphi_j(y)) \leq \epsilon_0 \text{ for } j=0,\ldots,n-1. \]

So (see \( \S 9 \)) if \( \epsilon_0 \) is small enough, then the unstable distance from \( x \) to \( y \) is \( \leq \frac{C \epsilon_0}{\lambda^n} \).
That is, each element \( A \in \mathcal{F} \) is contained in an unstable rectangle:

Thus \( \mu(A) \leq \frac{C^2 \varepsilon^2}{1 / \lambda^n} \).

Recalling the formula for the entropy \( H_\mu(\mathcal{F}(\varepsilon)) \), we see that

\[
H_\mu(\mathcal{F}(\varepsilon)) \geq -\log \frac{C^2 \varepsilon^2}{1 / \lambda^n} = n \log |\lambda| + O(1) \quad \text{as } n \to \infty
\]

So \( h_\mu(\varepsilon, \mathcal{F}) \geq \log |\lambda| \).

§10.3. Generating partitions

It turns out that \( h_\mu(\varepsilon) = h_\mu(\varepsilon, \mathcal{F}) \) for a sufficiently fine partition \( \mathcal{F} \) in many cases. Denote by \( \mathcal{D}_m \) the set of all partitions of \( X \) into \( m \) sets.
We define the following metric on $P_{\mu}$:

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

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

**Definition:** If $\psi : \Xi \rightarrow \Xi$ is a map preserving a probability measure $\mu$ and $\xi$ is a partition, we say that $\xi$ is:
- a *one-sided generator* for $\psi$, if
  - $\forall \eta$, $\forall \delta > 0$
  - $\exists \eta$, $\forall \xi \subseteq \xi^{(n)}$
  - (here $\xi^{(n)} = \xi \psi^{-1}(\xi) \psi \ldots \psi^{-1}(\xi)$)
  - such that $d(\eta, \xi) \leq \delta$.

That is, any finite partition is well-approximated by partitions subordinate to $\xi^{(n)}$. 
- If \( \varphi \) is invertible, we call \( \xi \) a generator for \( \varphi \), if the same property holds with \( \xi^{(n)} \) replaced by 
\[
\bigvee_{j=-n}^{n} \varphi_j(\xi).
\]

We will show

Thus assume that \( \xi \) is a one-sided generator, or \( \varphi \) is invertible & \( \xi \) is a generator.

Then \[
\mu(\varphi) = \mu(\varphi, \xi).
\]

Example: Assume that \( \mu \) is nonatomic and 
\[
\max_{AE\xi^{(n)}} \text{diam}(A) \to 0 \text{ as } n \to \infty.
\]

Then \( \xi \) is a one-sided generator.

(Similarly for \( \varphi \) invertible & \( \xi \) a generator)
Will skip the proof but this works similarly to approximating arbitrary Lebesgue measurable sets by unions of small cubes:

\[ \mathbb{S}^1 \]

So e.g.

- for irrational shift on \( \mathbb{S}^1 \) (example (1)),
  \[ \exists = [0, \frac{1}{2}), [\frac{1}{2}, 1] \text{ is a 1-sided generator} \]
  (so \( h_\mu(\psi) = 0 \))

- for the map \( x \mapsto 2x \) on \( \mathbb{S}^1 \),
  the same \( \exists \) is a 1-sided generator
  (so \( h_\mu(\psi) = \log 2 \))

- for the cat map, any \( \exists \) consisting of sets with small diameter is a generator (but it is not a 1-sided generator)
We now start the proof of Theorem with
\[ H_\mu ( \xi | \xi ) \]

**Lemma.** We have partitions \( \xi \) and \( \eta \)
\[ h_\mu ( \phi, \eta ) = h_\mu ( \phi, \xi ) + H_\mu ( \eta | \xi ). \]

**Proof.**

1. Denote
   \[ \xi^{(n)} := \xi \cup \varphi^{-1}(\xi) \cup \cdots \cup \varphi^{-(n-1)}(\xi), \]
   \[ \eta^{(n)} := \eta \cup \varphi^{-1}(\eta) \cup \cdots \cup \varphi^{-(n-1)}(\eta). \]
   Then
   \[ H_\mu ( \xi^{(n)} ) \leq H_\mu ( \xi^{(n)} \cup \eta^{(n)} ) = \]
   \[ = H_\mu ( \eta^{(n)} ) + H_\mu ( \xi^{(n)} | \eta^{(n)} ). \]

Recalling the definition of \( H_\mu \), we see that it suffices to prove the inequality
\[ (*) \quad H_\mu ( \xi^{(n)} | \eta^{(n)} ) \leq H( \xi | \eta ). \]

2. We show (*) by induction on \( n \).
   - \( n = 1 \) is immediate.
   - For \( n \geq 2 \), we use the identity (valid for any 3 partitions \( \xi, \eta, \xi \))
\[ (** ) \quad H_\mu (\xi V \eta | \zeta) = H_\mu (\xi | \zeta) + H_\mu (\eta | \zeta V \xi). \]

To check (**), we compute

\[
H_\mu (\xi V \eta | \zeta) = \sum_{A \in \mathcal{S}, B \in \mathcal{V}, C \in \mathcal{S}} \mu(\mathcal{A} \cap \mathcal{V} \cap \mathcal{C}) \log \frac{\mu(\mathcal{A} \cap \mathcal{V} \cap \mathcal{C})}{\mu(\mathcal{C})}.
\]

\[
= \sum_{A, B, C} \mu(\mathcal{A} \cap \mathcal{V} \cap \mathcal{C}) \log \left( \frac{\mu(\mathcal{A} \cap \mathcal{V} \cap \mathcal{C})}{\mu(\mathcal{A} \cap \mathcal{C})} \cdot \frac{\mu(\mathcal{A} \cap \mathcal{C})}{\mu(\mathcal{C})} \right) + H(\eta | \zeta V \xi),
\]

\[
= \sum_{A, B, C} \mu(\mathcal{A} \cap \mathcal{V} \cap \mathcal{C}) \log \frac{\mu(\mathcal{A} \cap \mathcal{V} \cap \mathcal{C})}{\mu(\mathcal{A} \cap \mathcal{C})} + H(\eta | \zeta V \xi),
\]

\[
= H(\xi | \eta) + H(\varphi^{-1}(\xi^{(n-1)}) | \xi V \eta^{(n)}).
\]

\[
= H(\xi | \eta) + H(\varphi^{-1}(\xi^{(n-1)}) | \varphi^{-1}(\eta^{(n)})),
\]

(\text{as } \eta \in \eta^{(n)}, \ \varphi^{-1}(\eta^{(n-1)}) \leq \xi V \eta^{(n)}).

\[
\leq H(\xi | \eta) + H(\xi^{(n-1)} | \eta^{(n-1)}),
\]

as \( \varphi \) is measure preserving.

\[
\Rightarrow H(\xi | \eta) + H(\xi^{(n-1)} | \eta^{(n-1)}).
\]

\[ \Rightarrow \text{by induction} \quad \text{on } n. \]
Corollary: if \( \eta \leq \xi \) then
\[
H_{\mu}(\eta, \xi) = 0 \quad \text{and thus} \quad h_{\mu}(\varphi, \eta) \leq h_{\mu}(\varphi, \xi).
\]

Lemma: We have \( \forall m \geq 0 \)
\[
h_{\mu}(\varphi, \xi^{(m)}) = h_{\mu}(\varphi, \xi).
\]

If \( \varphi \) is invertible, same is true with \( \xi^{(m)} \) replaced by \( \bigcap_{l=-m}^{m} \varphi^{l}(\xi) \).

Proof: We just show the first statement: the second one holds since
\[
\bigcap_{l=-m}^{m} \varphi^{l}(\xi) = \varphi^{m}(\xi^{(2m+1)}).
\]

The \( n \)-th refinement of \( \xi^{(m)} \) is
\[
(\xi^{(m)})^{(n)} = \bigcap_{j=0}^{n-1} \bigcap_{l=0}^{m-1} \varphi^{-j}(\xi^{(m)}) = \bigcap_{l=0}^{m-n-1} \varphi^{-j}(\xi^{m+n})
\]

Then
\[
h_{\mu}(\varphi, \xi^{(m)}) = \lim_{n \to \infty} H_{\mu}(\xi^{(m+n)}) = \lim_{n \to \infty} h_{\mu}(\xi^{(m+n)}) = h_{\mu}(\varphi, \xi). \]
Lemma. Fix $m \geq 1$ and let $P_m$ be the set of partitions with $m$ elements.

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

$\forall \xi, \eta \in P_m$, if $d(\xi, \eta) \leq \delta$ then $H_\mu(\eta | \xi) \leq \varepsilon$.

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

Proof. 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 $\alpha_j := \frac{\mu(A_j \setminus B_j)}{\mu(A_j)} = \mu(B_j | A_j)$, 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

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

and $\mu(B_j | A_j) = 1 - \alpha_j$.
We split the sum into 2 parts:

\[ j=k: \quad \text{get } -\sum_{j=1}^{m} \mu(A_j \cap B_j) \log \mu(B_j | A_j) = \]

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

\[ j \neq k: \quad \text{get } -\sum_{j=1}^{m} \mu(A_j) \sum_{k \neq j} \mu(B_k | A_j) \log \mu(B_k | A_j) \]

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

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

where \( c_{jk} = \frac{\mu(B_k | A_j)}{\alpha_j} = \frac{\mu(B_k \cap A_j)}{\mu(A_j \setminus B_j)} \)

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

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

\[ \leq -\sum_{j=1}^{m} \mu(A_j) \alpha_j (-\log \alpha_j + \log (m-1)) \]
Putting these together, we get $H_{\mu}(\eta|\xi) \leq 18.118$

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

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

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

Note that $\beta_j \leq \log m$, as

$$\beta_j = (1-d_j)\log \frac{1}{1-d_j} + d_j\log \frac{m-1}{d_j} \leq \log(1-d_j \cdot \frac{1}{1-d_j} + d_j \cdot \frac{m-1}{d_j}) = \log m.$$  

Put $\Psi(x) := -(1-x)\log(1-x) - x\log x$, then $\Psi$ is increasing on $[0, \frac{1}{2}]$.

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

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

$$= I + II.$$

If $\mu(A_j) > \sqrt{\delta}$, then, since $\sum_j \mu(A_j) d_j \leq \delta$, we have $d_j \leq \sqrt{\delta} \Rightarrow \Psi(d_j) \leq \Psi(\sqrt{\delta}).$
Thus
\[
I = \sum_{\mu(A_j) \geq \sqrt{\delta}} \mu(A_j) \cdot (4(d_j) + d_j \log(n-1))
\]
\[
\leq 4(\sqrt{\delta}) + \sqrt{\delta} \log(n-1)
\]
And
\[
II = \sum_{\mu(A_j) \leq \sqrt{\delta}} \mu(A_j) \log n \leq \sqrt{\delta} \cdot m \log m.
\]
So
\[
H_\mu(\eta|\xi) \leq 4(\sqrt{\delta}) + \sqrt{\delta} \cdot (m \log m + \log(n-1))
\]
which goes to 0 as \( \delta \to 0 \).

\[\square\]

We can now give

**Proof of Theorem**

We assume \( \xi \) is a one-sided generator

(the case of \( \eta \) invertible & \( \xi \) a generator

is handled similarly).

Since
\[
H_\mu(\eta) = \sup \{ H_\mu(\phi, \eta) : \eta \text{ a finite partition} \},
\]

to show that
\[
H_\mu(\eta) = H_\mu(\phi, \xi)
\]

it suffices to prove that A finite partition \( \eta \)

we have

\[
H_\mu(\phi, \eta) \leq H_\mu(\phi, \xi).
\]

Put \( m := \text{number of elements in } \eta \).
Take arbitrary $\varepsilon > 0$. Let $\delta > 0$ be from the last lemma.

Since $\xi$ is a one-sided generator,

$\exists \eta_0$ and a partition $\xi \leq \xi^{(n)}$

such that $d(\eta, \xi) \leq \delta$.

By the last lemma, we have

$H_\mu(\eta | \xi) \leq 3$.

Now,

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

(as $\xi \leq \xi^{(n)}$)

$\leq h_\mu(\varphi, \xi^{(n)}) + 3$

$\leq h_\mu(\varphi, \xi) + 3$

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

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

as needed.