Entropy of Measure Preserving Endomorphisms

Haiyu Zhang

March 15, 2024

Partition

Let $(X, \mathcal{F}, \mu)$ be a probability space.

Definition

A **partition** of $(X, \mathcal{F}, \mu)$ is a subfamily (a priori may be uncountable) of $\mathcal{F}$ consisting of mutually disjoint elements whose union is X.

Given a partition denoted by $\mathcal{A}$ and $x \in X$, the only element of $\mathcal{A}$ contaning x is denoted by $\mathcal{A}(x)$ or, if $x \in A \in \mathcal{A}$, by $A(x)$.
Suppose $\mathcal{A}$ and $\mathcal{B}$ are two partitions of X,

- We define their *join* or *joining*

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

- We say $\mathcal{B}$ is *finer* than $\mathcal{A}$ or that $\mathcal{B}$ is a *refinement* of $\mathcal{A}$, denoted by $\mathcal{A} \leq \mathcal{B}$, if and only if $\mathcal{B}(x) \subset \mathcal{A}(x)$ for every $x \in X$.In this case, $\mathcal{A} \vee \mathcal{B} = \mathcal{B}$.

Entropy of a Countable Partition

Definition

Let $\mathcal{A} = \{A_i : 1 \leq i \leq n\}$ be a countable partition of X, where $n \geq 1$ is a finite integer or ∞. The **entropy** of $\mathcal{A}$ is the number

$$\mathrm{H}(\mathcal{A}) = \sum_{i=1}^{\infty} -\mu(A_i) \log \mu(A_i) = \sum_{i=1}^{\infty} k(\mu(A_i))$$

where the function $k : [0,1] \to [0,\infty]$ is defined by

$$k(t) = \begin{cases} -t \log t & \text{for } t \in (0,1] \\ 0 & \text{for } t = 0 \end{cases}$$

Note that k is a concave function with two following properties:

- $k(\lambda_1 x_1 +_2 x_2) \geq \lambda_1 k(x_1) + \lambda_2 k(x_2)$ $(\lambda_1, \lambda_2 > 0 \ \lambda_1 + \lambda_2 = 1)$
- $k(x_1 + x_2) \leq k(x_1) + k(x_2)$

Remark 1:

If $\mathcal{A}$ is finite, say consists of n elements, then $0 \leq \mathrm{H}(\mathcal{A}) \leq \log n$.

- $\mathrm{H}(\mathcal{A}) = 0$ for $\mathcal{A} = \{X\}$
- $\mathrm{H}(\mathcal{A}) = \log n$ if and only if $\mu(A_1) = \mu(A_2) = \ldots = \mu(A_n) = 1/n$.

Remark 2:

$$H(A) = \int -\log \mu(\mathcal{A}(x))\, d\mu$$

where $I(x) = I(\mathcal{A})(x) := -\log \mu(\mathcal{A}(x))$ is called an *information function*.

Remark 3:

$$\mathrm{H}(\mathcal{A} \vee \mathcal{B}) = \sum_{i,j} -\mu(A_i \cap B_j) \log \mu(A_i \cap B_j)$$

is called *joint entropy*.

Conditional Entropy

Definition

Let $\mathcal{A} = \{A_i : i \geq 1\}$ and $\mathcal{B} = \{B_j : j \geq 1\}$ be two countable partitions of X. The **conditional entropy** $\mathrm{H}(\mathcal{A} \mid \mathcal{B})$ of $\mathcal{A}$ given $\mathcal{B}$ is defined as

$$\begin{aligned}\mathrm{H}(\mathcal{A} \mid \mathcal{B}) &= \sum_{j=1}^{\infty} \mu(B_j) \sum_{i=1}^{\infty} -\frac{\mu(A_i \cap B_j)}{\mu(B_j)} \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \\ &= \sum_{i,j} -\mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)}\end{aligned}$$

Define $I(\mathcal{A}|\mathcal{B})(x) := -\log \mu(A(x) \mid B(x)) = -\log \frac{\mu(A(x) \cap B(x))}{\mu(B(x))}\Big)$,
then

$$\mathrm{H}(\mathcal{A} \mid \mathcal{B}) = \int_X I(\mathcal{A} \mid \mathcal{B}) d\mu.$$

If $\tilde{\mathcal{B}}$ is the σ-algebra consisting of all unions of elements of $\mathcal{B}$ (i.e. generated by $\mathcal{B}$), then
$I(x) = -\log \mu((\mathcal{A}(x) \cap \mathcal{B}(x)) \mid \mathcal{B}(x)) = -\log E\left(\not\Vdash_{\mathcal{A}(x)} \mid \tilde{\mathcal{B}}\right)(x)$

Properties of Conditional Entropy

Theorem

Let $(X, \mathcal{F}, \mu)$ be a probability space. If $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ are countable partitions of X then:

$$\mathrm{H}(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = \mathrm{H}(\mathcal{A} \mid \mathcal{C}) + \mathrm{H}(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C})$$
$$\mathrm{H}(\mathcal{A} \vee \mathcal{B}) = \mathrm{H}(\mathcal{A}) + \mathrm{H}(\mathcal{B} \mid \mathcal{A})$$
$$\mathcal{A} \leq \mathcal{B} \Rightarrow \mathrm{H}(\mathcal{A} \mid \mathcal{C}) \leq \mathrm{H}(\mathcal{B} \mid \mathcal{C})$$
$$\mathcal{B} \leq \mathcal{C} \Rightarrow \mathrm{H}(\mathcal{A} \mid \mathcal{B}) \geq \mathrm{H}(\mathcal{A} \mid \mathcal{C})$$
$$\mathrm{H}(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) \leq \mathrm{H}(\mathcal{A} \mid \mathcal{C}) + \mathrm{H}(\mathcal{B} \mid \mathcal{C})$$
$$\mathrm{H}(\mathcal{A} \mid \mathcal{C}) \leq \mathrm{H}(\mathcal{A} \mid \mathcal{B}) + \mathrm{H}(\mathcal{B} \mid \mathcal{C})$$

Proof.

Let $\mathcal{A} = \{A_n : n \geq 1\}, \mathcal{B} = \{B_m : m \geq 1\}$, and $\mathcal{C} = \{C_l : l \geq 1\}$. WLOG. we assume that all these sets are of positive measure.

(a) By definition,

$$\mathrm{H}(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = -\sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(C_k)}$$

Suppose $\mu(A_i \cap C_k) \neq 0$ (note that if it equals 0, left hand side will vanish and hence we need not consider it), we have

$$\frac{\mu(A_i \cap B_j \cap C_k)}{\mu(C_k)} = \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(A_i \cap C_k)} \frac{\mu(A_i \cap C_k)}{\mu(C_k)}$$

Therefore,

$$\begin{aligned}
\mathrm{H}(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = & -\sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \\
& -\sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(A_i \cap C_k)} \\
= & -\sum_{i,k} \mu(A_i \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} + \mathrm{H}(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C}) \\
= & \mathrm{H}(\mathcal{A} \mid \mathcal{C}) + \mathrm{H}(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C})
\end{aligned}$$

(b) Put $\mathcal{C} = \{X\}$ in (a).
(c) By (a)

$$\mathrm{H}(\mathcal{B} \mid \mathcal{C}) = \mathrm{H}(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = \mathrm{H}(\mathcal{A} \mid \mathcal{C}) + \mathrm{H}(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C}) \geq \mathrm{H}(\mathcal{A} \mid \mathcal{C})$$

(d) Since the function k defined in the definition of entropy is strictly concave, we have for every pair i, j that

$$k\left(\sum_l \frac{\mu\left(C_l \cap B_j\right)}{\mu\left(B_j\right)} \frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)}\right) \geq \sum_l \frac{\mu\left(C_l \cap B_j\right)}{\mu\left(B_j\right)} k\left(\frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)}\right)$$

Note that $C_l \cap B_j = C_l$ since $\mathcal{B} \leq \mathcal{C}$,we get

$$k\left(\sum_l \frac{\mu\left(A_i \cap C_l\right)}{\mu\left(B_j\right)}\right) = k\left(\frac{\mu\left(A_i \cap B_j\right)}{\mu\left(B_j\right)}\right) \geq \sum_l \frac{\mu\left(C_l \cap B_j\right)}{\mu\left(B_j\right)} k\left(\frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)}\right).$$

Finally, multiplying both sides of (1.3.7) by $\mu\left(B_j\right)$ using the definition of k and summing over i and j we get

$$
\begin{aligned}
& -\sum_{i,j} \mu\left(A_i \cap B_j\right) \log \frac{\mu\left(A_i \cap B_j\right)}{\mu\left(B_j\right)} \\
\geq & -\sum_{i,j,l} \mu\left(C_l \cap B_j\right) \frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)} \log \frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)} \\
= & -\sum_{i,l} \mu\left(C_l\right) \frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)} \log \frac{\mu\left(A_i \cap C_l\right)}{\mu\left(C_l\right)}
\end{aligned}
$$

By definition, $\mathrm{H}(\mathcal{A} \mid \mathcal{B}) \geq \mathrm{H}(\mathcal{A} \mid \mathcal{C})$.
(e) We get $\mathrm{H}(\mathcal{B} \mid \mathcal{C}) \geq \mathrm{H}(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C})$ by (d) and it follows immediately from (a)
(f) By (d), (a) and (c)

$$
RHS \geq \mathrm{H}(\mathcal{A} \mid \mathcal{B} \vee \mathcal{C}) + \mathrm{H}(\mathcal{B} \mid \mathcal{C}) = \mathrm{H}(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) \geq \mathrm{H}(\mathcal{A} \mid \mathcal{C}) = LHS
$$

Entropy of Endomorphism Respect to a Partition

Let $(X, \mathcal{F}, \mu)$ be a probability space and let $T : X \to X$ be a measure preserving endomorphism of X.
If $\mathcal{A} = \{A_i\}_{i \in I}$ is a partition of X then by $T^{-1}\mathcal{A}$ we denote the partition $\left\{T^{-1}(A_i)\right\}_{i \in I}$.
For all $n \geq m \geq 0$ denote the partition $\bigvee_{i=0}^{n} T^{-i}\mathcal{A} = \mathcal{A} \vee T^{-1}(\mathcal{A}) \vee \cdots \vee T^{-n}(\mathcal{A})$ by $\mathcal{A}^n$ and $\bigvee_{i=m}^{n} T^{-i}(\mathcal{A})$ by A_m^n.

Remark 1:

Note that for any countable $\mathcal{A}$

$$\mathrm{H}\left(T^{-1}\mathcal{A}\right) = \mathrm{H}(\mathcal{A})$$

Remark 2:

If A and B are two subsets of X, $T^{-1}(A \cap B) = T^{-1}(A) \cap T^{-1}(B)$, hence $\mathcal{A}_k^{k+n} = T^{-k}(\mathcal{A}^n)$ and

$$\mathrm{H}\left(\mathcal{A}_k^{k+n}\right) = \mathrm{H}(\mathcal{A}^n)$$

Lemma

For any countable partition $\mathcal{A}$,

$$\mathrm{H}\left(\mathcal{A}^{n}\right)=\mathrm{H}(\mathcal{A})+\sum_{j=1}^{n} \mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_{1}^{j}\right)$$

Proof.

We prove this formula by induction. If $n=0$, it is tautology. Suppose it is true for $n-1 \geq 0$. Then with the use of properties of conditional entropy (b) and the measure preserving property of T we obtain

$$\begin{aligned} \mathrm{H}\left(\mathcal{A}^{n}\right) & =\mathrm{H}\left(\mathcal{A}_{1}^{n} \vee \mathcal{A}\right)=\mathrm{H}\left(\mathcal{A}_{1}^{n}\right)+\mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_{1}^{n}\right) \\ & =\mathrm{H}\left(\mathcal{A}^{n-1}\right)+\mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_{1}^{n}\right)=\mathrm{H}(\mathcal{A})+\sum_{j=1}^{n} \mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_{1}^{j}\right) \end{aligned}$$

By induction, it holds for all n.

Lemma

The sequences $\frac{1}{n+1}\mathrm{H}\left(\mathcal{A}^n\right)$ *and* $\mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_1^n\right)$ *are monotone decreasing to a limit* $\mathrm{h}(T, \mathcal{A})$*.*

Proof.

The sequence $\mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_1^n\right), n = 0, 1, \dots$ is monotone decreasing by properties of conditional entropy (d). By the monotone bounded convergence theorem, $\frac{1}{n+1}\sum_{j=1}^{n} \mathrm{H}\left(\mathcal{A} \mid \mathcal{A}_1^j\right)$ is also monotone decreasing to the same limit, furthermore it coincides with the limit of the sequence $\frac{1}{n+1}\mathrm{H}\left(\mathcal{A}^n\right)$ by the first lemma.

Definition

Given a probability space $(X, \mathcal{F}, \mu)$, a partition $\mathcal{A}$ on X and a measure preserving endomorphism $T : X \to X$, the limit $\frac{1}{n+1}\mathrm{H}\left(\mathcal{A}^n\right)$ is called **(measure-theoretic) entropy** of T with respect to the partition $\mathcal{A}$ which is denoted by $\mathrm{h}(T, \mathcal{A})$ or by $\mathrm{h}_\mu(T, \mathcal{A})$.

The lemma above shows the existence of it. Intuitively this means the limit rate of the growth of average (integral) information (in logarithmic scale), under consecutive experiments, for the number of those experiments tending to infinity.

Properties of $\mathrm{h}(T, \mathcal{A})$

Theorem

If $\mathcal{A}$ and $\mathcal{B}$ are countable partitions of finite entropy then

$$\mathrm{h}(T, \mathcal{A}) \leq \mathrm{H}(\mathcal{A})$$
$$\mathrm{h}(T, \mathcal{A} \vee \mathcal{B}) \leq \mathrm{h}(T, \mathcal{A}) + \mathrm{h}(T, \mathcal{B})$$
$$\mathcal{A} \leq \mathcal{B} \Rightarrow \mathrm{h}(T, \mathcal{A}) \leq \mathrm{h}(T, \mathcal{B})$$
$$\mathrm{h}(T, \mathcal{A}) \leq \mathrm{h}(T, \mathcal{B}) + \mathrm{H}(\mathcal{A} \mid \mathcal{B})$$
$$\mathrm{h}\left(T, T^{-1}(\mathcal{A})\right) = \mathrm{h}(T, \mathcal{A})$$
$$\textit{If } k \geq 1 \textit{ then } \mathrm{h}(T, \mathcal{A}) = \mathrm{h}\left(T, \mathcal{A}^k\right)$$
$$\textit{If } T \textit{ is invertible and } k \geq 1, \textit{ then } \mathrm{h}(T, \mathcal{A}) = \mathrm{h}\left(T, \bigvee_{i=-k}^{k} T^i(\mathcal{A})\right)$$

Proof.
(a) By properties of conditional entropy (d), $\mathrm{H}(\mathcal{A}^j)$ is decreasing.
(b) can be proved by properties of conditional entropy (e).
(c) $\mathcal{A} \leq \mathcal{B} \Rightarrow \mathcal{A}^n \leq \mathcal{B}^n \Rightarrow \mathrm{h}(T, \mathcal{A}) \leq \mathrm{h}(T, \mathcal{B})$ by properties of conditional entropy (c).
(d)

$$\begin{aligned}
\mathrm{h}(T, \mathcal{A}) &= \lim_{n\to\infty} \frac{1}{n} \mathrm{H}\left(\mathcal{A}^{n-1}\right) \leq \lim \mathrm{H}(\mathcal{A}^{n-1} \vee \mathcal{B}^{n-1}) \\
&= \lim_{n\to\infty} \frac{1}{n} \left(\mathrm{H}\left(\mathcal{A}^{n-1} \mid \mathcal{B}^{n-1}\right) + \mathrm{H}\left(\mathcal{B}^{n-1}\right)\right) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} H\left(T^{-j}(\mathcal{A}) \mid \mathcal{B}^{n-1}\right) + \lim_{n\to\infty} \frac{1}{n} \mathrm{H}\left(\mathcal{B}^{n-1}\right) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} H\left(T^{-j}(\mathcal{A}) \mid T^{-j}(\mathcal{B})\right) + \mathrm{h}(T, \mathcal{B}) \\
&= \mathrm{H}(\mathcal{A} \mid \mathcal{B}) + \mathrm{h}(T, \mathcal{B}).
\end{aligned}$$

(e)

$$LHS = \lim_{n\to\infty} \frac{1}{n+1}\mathrm{H}(\mathcal{A}_1^{n+1}) = \lim_{n\to\infty} \frac{1}{n+1}\mathrm{H}(\mathcal{A}^n) = RHS$$

(f) By (c), $\mathrm{h}(T, \mathcal{A}) \leq \mathrm{h}\left(T, \mathcal{A}^k\right)$; Also note that

$$\mathrm{H}((\mathcal{A}^k)^n) = \mathrm{H}(\mathcal{A}^k \vee \mathcal{A}_{k+1}^{k+n}) \leq \mathrm{H}(\mathcal{A}^k) + \mathrm{H}(\mathcal{A}_{k+1}^{k+n+1})$$

Hence

$$\begin{aligned} \mathrm{h}\left(T, \mathcal{A}^k\right) &\leq \lim_{n\to\infty} \frac{1}{n+1}\mathrm{H}(\mathcal{A}^k) + \lim_{n\to\infty} \frac{1}{n+1}\mathrm{H}(\mathcal{A}_{k+1}^{k+n+1}) \\ &= \lim_{n\to\infty} \frac{1}{n+1}\mathrm{H}(\mathcal{A}^n) = \mathrm{h}(T, \mathcal{A}) \end{aligned}$$

(g) Use $T \circ T^{-1} = I$(identity map) and (f).

Theorem

If $T : X \to X$ is a measure preserving endomorphism of a probability space $(X, \mathcal{F}, \mu), \mathcal{A}$ and $\mathcal{B}_m, m = 1, 2, \ldots$ are countable partitions with finite entropy, and $\mathrm{H}(\mathcal{A} \mid \mathcal{B}_m) \to 0$ as $m \to \infty$, then

$$\mathrm{h}(T, \mathcal{A}) \leq \liminf_{m \to \infty} \mathrm{h}(T, \mathcal{B}_m).$$

In particular, for $\mathcal{B}_m := \mathcal{B}^m = \bigvee_{j=0}^{m} T^{-j}(\mathcal{B})$, one obtains $\mathrm{h}(T, \mathcal{A}) \leq \mathrm{h}(T, \mathcal{B})$.

Proof.

By property (d) above, we get for every positive integer m, that

$$\mathrm{h}(T, \mathcal{A}) \leq \mathrm{H}(\mathcal{A} \mid \mathcal{B}_m) + \mathrm{h}(T, \mathcal{B}_m).$$

Letting $m \to \infty$ this yields the first part of the assertion since $\mathrm{H}(\mathcal{A} \mid \mathcal{B}_m) \to 0$. If $\mathcal{B}_m = \mathcal{B}^m$, then $\mathrm{h}(T, \mathcal{B}^m) = \mathrm{h}(T, \mathcal{B})$, by property (f) above, and the second part of the theorem follows as well.

Remark: Note that this theorem is more stronger than the property (c) above since $\mathcal{A} \leq \mathcal{B} \Rightarrow \mathrm{H}(\mathcal{A} \mid \mathcal{B}_m) = 0$.

(Measure-theoretic) Entropy of an Endomorphism

Definition

The **(measure-theoretic) entropy of an endomorphism** $T : X \to X$ is defined as

$$\mathrm{h}_\mu(T) = \mathrm{h}(T) = \sup_{\mathcal{A}} \{\mathrm{h}(T, \mathcal{A})\}$$

where the supremum is taken over all finite (or countable of finite entropy) partitions of X.

Remark:

- From the definition we know that the entropy of T is an isomorphism invariant.
- By the theorem above, if $\mathrm{H}(\mathcal{A} \mid \mathcal{B}^m) \to 0$ for all partition $\mathcal{A}$, then $\mathrm{h}(T) = \mathrm{h}(T, \mathcal{B})$.

Properties of $\mathrm{h}_\mu(T)$

Theorem

If $T : X \to X$ is a measure preserving endomorphism of a probability space $(X, \mathcal{F}, \mu)$ then

$$\mathrm{h}\left(T^k\right) = k\ \mathrm{h}(T) \text{ for all } k \geq 1,$$

$$\text{If } T \text{ is invertible then } \mathrm{h}\left(T^{-1}\right) = \mathrm{h}(T).$$

Proof.
(a) Fix $k \geq 1$. We have

$$\begin{aligned}
\mathrm{h}\left(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\right) &= \lim_{n\to\infty} \frac{1}{n} \mathrm{H}\left(\bigvee_{j=0}^{n-1} T^{-kj}\left(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\right)\right) \\
&= \lim_{n\to\infty} \frac{k}{nk} \mathrm{H}\left(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{A}\right) = k\ \mathrm{h}(T, \mathcal{A})
\end{aligned}$$

Therefore,

$$
\begin{aligned}
k\ \mathrm{h}(T) = k \sup_{\mathcal{A}\text{ finite}} \mathrm{h}(T,\mathcal{A}) &= \sup_{\mathcal{A}} \mathrm{h}\left(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\right) \\
&\le \sup_{\mathcal{B}} \mathrm{h}\left(T^k, \mathcal{B}\right) = \mathrm{h}\left(T^k\right)
\end{aligned}
$$

On the other hand, by property (c) above, we get
$\mathrm{h}\left(T^k,\mathcal{A}\right) \le \mathrm{h}\left(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\right) = k\ \mathrm{h}(T,\mathcal{A})$
Therefore, $\mathrm{h}\left(T^k\right) = \sup_{\mathcal{B}} \mathrm{h}\left(T^k,\mathcal{B}\right) \le \sup_{\mathcal{B}} \mathrm{h}\left(T,\mathcal{B}\right) \le k\ \mathrm{h}(T)$.
Hence the first equation is proved.
(b) For all finite partitions $\mathcal{A}$ we have

$$
\mathrm{H}\left(\bigvee_{i=0}^{n-1} T^{i}\mathcal{A}\right) = \mathrm{H}\left(T^{-(n-1)} \bigvee_{i=0}^{n-1} T^{i}\mathcal{A}\right) = \mathrm{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right)
$$

This finishes the proof.

Theorem

If μ and ν are two probability measures on $(X, \mathcal{F})$ both preserved by an endomorphism $T : X \to X$. Then for every $a : 0 < a < 1$ and the measure $\rho = a\mu + (1-a)\nu$ we have

$$\mathrm{h}_\rho(T) = a\ \mathrm{h}_\mu(T) + (1-a)\ \mathrm{h}_\nu(T).$$

In other words the mapping $\mu \mapsto \mathrm{h}_\mu$ is affine.

Proof.

By the concavity of the function $k(t) = -t\log t$, for every $A \in \mathcal{F}$ we have
$0 \leq k(\rho(A)) - ak(\mu(A)) - (1-a)k(\nu(A))$
$\leq -(a\log a)\mu(A) - ((1-a)\log(1-a))\nu(A)$,
Summing it up over $A \in \mathcal{A}$ for a finite partition $\mathcal{A}$, we obtain

$$0 \leq \mathrm{H}_\rho(\mathcal{A}) - a\mathrm{H}_\mu(\mathcal{A}) - (1-a)\mathrm{H}_\nu(\mathcal{A}) \leq \log 2.$$

Hence, for any finite partiton $\mathcal{A}$,

$$\mathrm{h}_\mu(T, \mathcal{A}) = a\ \mathrm{h}_\mu(T, \mathcal{A}) + (1-a)\ \mathrm{h}_\nu(T, \mathcal{A})$$

and apply this to partitions $\mathcal{A}^n$,we obtain

$$\mathrm{h}_\rho(T) \leq \ \mathrm{h}_\mu(T) + (1-a)\ \mathrm{h}_\nu(T).$$

Let $\{\mathcal{A}_i\}, \{\mathcal{B}_i\} \subset \mathcal{F}$ be a set of partitions such that

$$\lim_{i\to\infty} \mathrm{h}_\mu(T, \mathcal{A}_i) = \mathrm{h}_\mu(T),\ \lim_{i\to\infty} \mathrm{h}_\nu(T, \mathcal{B}_i) = \mathrm{h}_\nu(T)$$

Consider a set of partitions $\{\mathcal{A}_i \vee \mathcal{B}_i\}$,

$$\begin{aligned} \mathrm{h}_\rho(T) \geq \lim_{i\to\infty}(T, \mathcal{A}_i \vee \mathcal{B}_i) &\geq a \lim_{i\to\infty} \mathrm{h}_\mu(T, \mathcal{A}_i) + (1-a) \lim_{i\to\infty} \mathrm{h}_\nu(T, \mathcal{B}_i) \\ &= a\mathrm{h}_\mu + (1-a)\mathrm{h}_\nu(T) \end{aligned}$$

Hence the equation is proved.