# Lecture 7

# 4 Entropy

In 1958, Kolomogrov introduced the notion of entropy in ergodic theory. The entropy of a measure-preserving transformation $T$ is defined is three stages: the entropy of a finite partition, the entropy of $T$ relative to a finite partition and the entropy of $T$.

## 4.1 Entropy of finite partitions

Let $(X, \mathscr{B}, m)$ be a probability space.

**Definition 4.1** (Partition)**.** *A partition of $X$ is a family of elements in $\mathscr{B}$ which are disjoint and the union of them is $X$.*

For instance, $\mathscr{N} = \{\emptyset, X\}$ is a partition, called the trivial partition.

**Definition 4.2** (Entropy of a finite partition)**.** *Let $\xi = \{A_1, \cdots, A_n\}$ be a partition of $X$. Define the entropy of $\xi$ by*

$$H(\xi) := -\sum_{i=1}^{n} m(A_i) \log m(A_i) = \sum_{i=1}^{n} \phi(m(A_i)),$$

*where $\phi(x) := -x \log x$ for $x \in (0, 1]$ and $\phi(0) := 0$.*

To study the properties of entropy, we will use the following two simple lemmas, which can be proved by elementary calculus.

**Lemma 4.1.** *The function $\phi : x \mapsto -x \log x$ is strictly concave on $(0, 1)$.*

**Lemma 4.2** (Jensen's inequality)**.** *Let $a_1, \cdots, a_n \in [0, 1]$ with $\sum_{i=1}^{n} a_i = 1$. Let $x_1, \cdots, x_n \in [0, 1]$. Then*

$$\sum_{i=1}^{n} a_i \phi(x_i) \leq \phi(\sum_{i=1}^{n} a_i x_i).$$

*Moreover, "$=$" holds iff $x_i = x_j$ whenever $a_i, a_j \neq 0$.*

In the following, we will use symbols "$\mathring{=}, \mathring{\leq}, \mathring{\subseteq}$" to denote the ignorance of sets of measure 0. For instance, $\xi \mathring{=} \eta$ iff for any element $A \in \xi$, there is an element $B \in \eta$, such that $m(A \triangle B) = 0$ and vise versa.

**Lemma 4.3.** *Let $\xi = \{A_1, \cdots, A_n\}$ be a partition of $X$. Then $0 \leq H(\xi) \leq \log n$. Moreover, $H(\xi) = 0$ iff $\xi \mathring{=} \{\emptyset, X\}$ and $H(\xi) = \log n$ iff $m(A_i) = \frac{1}{n}$ for all $i$.*

*Proof.* Since $\phi$ is nonnegative on $[0,1]$ and attains 0 only when $x=0$ or $x=1$, $H(\xi)\geq 0$. If $H(\xi)=0$, then for each $i$, $m(A_i)=0$ or $m(A_i)=1$, hence $\xi \overset{\circ}{=} \{\emptyset, X\}$. In Lemma 4.2, put $a_i=\frac{1}{n}$ for all $i$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\phi(m(A_i)) \leq \phi(\frac{m(A_1)+\cdots+m(A_n)}{n}) = \phi(\frac{1}{n}) = \frac{\log n}{n},$$

hence $H(\xi)\leq \log n$ and "=" holds iff $m(A_i)=\frac{1}{n}$ for all $i$. □

**Definition 4.3.** *Let $\xi=\{A_1,\cdots,A_n\}$ and $\eta=\{B_1,\cdots,B_k\}$ be two partitions of $X$. The join of $\xi$ and $\eta$ is defined as*

$$\xi\vee\eta=\{A_i\cap B_j : 1\leq i\leq n, 1\leq j\leq k\}.$$

Notice that $\xi\vee\eta$ is a new partition, it is finer than $\xi$ and $\eta$ in the sense that any $A_i$ or $B_j$ is a union of elements in $\xi\vee\eta$.

**Definition 4.4** (Refinement of a partition)**.** *Given two finite partitions $\xi$ and $\eta$, write $\xi\leq\eta$ if any element of $\xi$ is a union of elements of $\eta$ and say $\eta$ is a refinement of $\xi$.*

## 4.2 Conditional entropy

**Definition 4.5** (Conditional entropy)**.** *Let $\xi=\{A_1,\cdots,A_n\}$ and $\eta=\{B_1,\cdots,B_k\}$ be two partitions of $X$. Define*

$$H(\xi|\eta)=\sum_{j=1}^{k}m(B_j)\sum_{i=1}^{n}\phi(\frac{m(A_i\cap B_j)}{m(B_j)}).$$

Intuitively, entropy (conditional entropy) measures the average uncertainty of determining the location of a typical point in $X$ (if its location in another partition is known). It is helpful to think in this way in the following theorems.

**Proposition 4.1.** *Let $\xi$ and $\eta$ be two finite partitions of $X$. Then*
*(i) $0\leq H(\xi|\eta)\leq H(\xi)$.*
*(ii) $H(\xi|\eta)=0 \Leftrightarrow \xi\overset{\circ}{\leq}\eta$.*
*(iii) $H(\xi|\eta)=H(\xi) \Leftrightarrow \xi$ and $\eta$ are independent in the sense that $m(A_i\cap B_j)=m(A_i)m(B_j)$ for $A_i\in\xi$, $B_j\in\eta$.*

*Proof.* (i) $H(\xi|\eta)\geq 0$ is obvious. Furthermore,

$$\begin{aligned}H(\xi|\eta)&=\sum_j m(B_j)\sum_i\phi(\frac{m(A_i\cap B_j)}{m(B_j)})=\sum_i\sum_j m(B_j)\phi(\frac{m(A_i\cap B_j)}{m(B_j)})\\&\leq\sum_i\phi(\sum_j m(B_j)\frac{m(A_i\cap B_j)}{m(B_j)})=\sum_i\phi(m(A_i))=H(\xi).\end{aligned}$$

(ii) Suppose $H(\xi|\eta) = 0$. Then for any $i, j$, we have $m(B_j)\phi(\frac{m(A_i \cap B_j)}{m(B_j)}) = 0$. If $m(B_j) > 0$, then $\frac{m(A_i \cap B_j)}{m(B_j)} = 0$ or 1, since $\sum_{i=1}^{n} \frac{m(A_i \cap B_j)}{m(B_j)} = 1$, there exists a unique $i$ such that $\frac{m(A_i \cap B_j)}{m(B_j)} = 1$ and $m(A_i \cap B_j) = 0$ for any other $i$, hence $\xi \overset{\circ}{\leq} \eta$. Conversely, if $\xi \overset{\circ}{\leq} \eta$, the for any $i, j$, if $m(A_i \cap B_j) \neq 0$, then $B_j \overset{\circ}{\subseteq} A_i$, so $\frac{m(A_i \cap B_j)}{m(B_j)} = 1$, hence every term in the summation of $H(\xi|\eta)$ is 0, that is $H(\xi|\eta) = 0$.

(iii) "$\Leftarrow$" is obvious. If $H(\xi|\eta) = H(\xi)$, then for each $i$,

$$\sum_j m(B_j)\phi(\frac{m(A_i \cap B_j)}{m(B_j)}) = \phi(\sum_j m(B_j)\frac{m(A_i \cap B_j)}{m(B_j)}),$$

hence $\frac{m(A_i \cap B_j)}{m(B_j)}$ is constant for those $j$ with $m(B_j) > 0$. Write $t_i = \frac{m(A_i \cap B_j)}{m(B_j)}$ for $j$ with $m(B_j) > 0$, then

$$m(A_i) = \sum_{j:m(B_j)>0} m(A_i \cap B_j) = \sum_{j:m(B_j)>0} m(B_j)t_i = t_i.$$

This completes the proof. □

**Theorem 4.4.** *Let $\xi$, $\eta$ and $\gamma$ be finite partitions of $X$. Then*

*(i) $H(\xi \vee \eta) = H(\xi) + H(\eta|\xi)$.*

*(ii) More generally, $H(\xi \vee \eta|\gamma) = H(\xi|\gamma) + H(\eta|\xi \vee \gamma)$.*

*Proof.* Notice that (i) follows from (ii) by replacing $\gamma$ by $\{\emptyset, X\}$, so it suffices to prove (ii). Let $\xi = \{A_i\}_i$, $\eta = \{B_j\}_j$ and $\gamma = \{C_k\}_k$. Then

$$\begin{aligned}
H(\xi \vee \eta|\gamma) &= \sum_k m(C_k) \sum_i \sum_j \phi\big(\frac{m(A_i \cap B_j \cap C_k)}{m(C_k)}\big) \\
&= -\sum_k m(C_k) \sum_{i,j} \frac{m(A_i \cap B_j \cap C_k)}{m(C_k)} \log \frac{m(A_i \cap B_j \cap C_k)}{m(C_k)} \\
&= -\sum_{i,j,k} m(A_i \cap B_j \cap C_k) \log \frac{m(A_i \cap B_j \cap C_k)}{m(C_k)} \\
&= -\sum_{i,j,k} m(A_i \cap B_j \cap C_k)\big[\log \frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)} + \log \frac{m(A_i \cap C_k)}{m(C_k)}\big] \\
&= -\sum_{i,j,k} m(A_i \cap B_j \cap C_k) \log \frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)} - \sum_{i,j,k} m(A_i \cap B_j \cap C_k) \log \frac{m(A_i \cap C_k)}{m(C_k)} \\
&= -\sum_{i,j,k} m(A_i \cap C_k) \cdot \frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)} \log \frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)} - \sum_{i,k} m(A_i \cap C_k) \log \frac{m(A_i \cap C_k)}{m(C_k)} \\
&= -\sum_{i,k} m(A_i \cap C_k) \sum_j \frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)} \log \frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)} \\
&\quad - \sum_k m(C_k) \sum_i \frac{m(A_i \cap C_k)}{m(C_k)} \log \frac{m(A_i \cap C_k)}{m(C_k)} \\
&= \sum_{i,k} m(A_i \cap C_k) \sum_j \phi\big(\frac{m(A_i \cap B_j \cap C_k)}{m(A_i \cap C_k)}\big) + \sum_k m(C_k) \sum_i \phi\big(\frac{m(A_i \cap C_k)}{m(C_k)}\big) \\
&= H(\eta|\xi \vee \gamma) + H(\xi|\gamma).
\end{aligned}$$

□

**Corollary 4.4.1.** *$H(\xi \vee \eta) \leq H(\xi) + H(\eta)$.*

*Proof.* By the above theorem, $H(\xi \vee \eta) = H(\xi) + H(\eta|\xi) \leq H(\xi) + H(\eta)$. □

**Theorem 4.5.** *Assume $\eta \leq \gamma$. Then $H(\xi|\gamma) \leq H(\xi|\eta)$.*

*Proof.* Since $\eta \leq \gamma$, we have $\gamma = \eta \vee \gamma$. Hence

$$\begin{aligned}
H(\xi|\gamma) = H(\xi|\eta \vee \gamma) &= \sum_{j,k} m(B_j \cap C_k) \sum_i \phi\big(\frac{m(A_i \cap B_j \cap C_k)}{m(B_j \cap C_k)}\big) \\
&= \sum_j m(B_j)\big[\sum_i \sum_k \frac{m(B_j \cap C_k)}{m(B_j)} \phi\big(\frac{m(A_i \cap B_j \cap C_k)}{m(B_j \cap C_k)}\big)\big] \\
&\leq \sum_j m(B_j)\big[\sum_i \phi\big(\sum_k \frac{m(B_j \cap C_k)}{m(B_j)} \times \frac{m(A_i \cap B_j \cap C_k)}{m(B_j \cap C_k)}\big)\big] \\
&= \sum_j m(B_j) \sum_i \phi\big(\frac{m(A_i \cap B_j)}{m(B_j)}\big) = H(\xi|\eta).
\end{aligned}$$

□

## 4.3 Entropy of $T$

First we define the entropy of a measure-preserving transformation $T$ w.r.t a partition $\xi$.

**Definition 4.6.** *Let $(X, \mathscr{B}, m, T)$ be a MPS. Let $\xi = \{A_1, \cdots, A_k\}$ be a a partition of $X$. Define*

$$h(T, \xi) := \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi),$$

*where $T^{-i}\xi := \{T^{-i}A_1, \cdots, T^{-i}A_k\}$ and $\bigvee_{i=0}^{n-1} T^{-i}\xi = \xi \bigvee T^{-1}\xi \bigvee \cdots \bigvee T^{-(n-1)}\xi$. We call $h(T, \xi)$ the entropy of $T$ w.r.t $\xi$.*

The existence of the above limit is guaranteed as follows.

**Lemma 4.6.** *$H(T^{-1}\xi) = H(\xi)$.*

*Proof.* It just follows from that $T$ preserves $m$. □

**Lemma 4.7.** *$H(\bigvee_{i=0}^{n+m-1} T^{-i}\xi) \leq H(\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(\bigvee_{i=0}^{m-1} T^{-i}\xi)$.*

*Proof.* Notice that

$$\begin{aligned} H(\bigvee_{i=0}^{n+m-1} T^{-i}\xi) &= H(\bigvee_{i=0}^{n-1} T^{-i}\xi \bigvee \bigvee_{i=n}^{n+m-1} T^{-i}\xi) = H(\bigvee_{i=0}^{n-1} T^{-i}\xi \bigvee T^{-n}(\bigvee_{i=0}^{m-1} T^{-i}\xi)) \\ &\leq H(\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(T^{-n}(\bigvee_{i=0}^{m-1} T^{-i}\xi)) = H(\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(\bigvee_{i=0}^{m-1} T^{-i}\xi). \end{aligned}$$

□

This lemma shows that $\{H(\bigvee_{i=0}^{n-1} T^{-i}\xi)\}_n$ is a subadditive sequence. Then the limit in Definition 4.6 exists by the following lemma.

**Lemma 4.8.** *Let $\{a_n\}$ be a sequence of real numbers. If $a_{n+m} \leq a_n + a_m$ for any $n, m \in \mathbb{N}_+$, then*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n}.$$

*Proof.* Fix $m \geq 1$. For $n \in \mathbb{N}_+$, there exist $q$ and $l$ such that $n = mq + l$, with $q \in \mathbb{N}$ and $0 \leq l < m$. Then

$$\frac{a_n}{n} = \frac{a_{mq+l}}{mq+l} \leq \frac{a_{mq} + a_l}{mq} \leq \frac{a_m}{m} + \frac{a_l}{mq},$$

letting $n \to \infty$, then $q \to \infty$, we have $\overline{\lim}_{n\to\infty} \frac{a_n}{n} \leq \frac{a_m}{m}$. Since $m$ is arbitrary, we have $\overline{\lim}_{n\to\infty} \frac{a_n}{n} \leq \inf_n \frac{a_n}{n}$. Since $\inf_n \frac{a_n}{n} \leq \underline{\lim}_{n\to\infty} \frac{a_n}{n}$, we complete the proof. Notice that the limit may be $-\infty$. □

Now we define the entropy of a measure-preserving transformation.

**Definition 4.7** (Entropy of $T$)**.** *Let $(X, \mathscr{B}, m, T)$ be MPS. Define*

$$h(T) = \sup_{\xi \text{ finite}} h(T, \xi).$$