Teaching Assistant Work (Ergodic Theory)

Xiao-Chuan Liu

November 2013

1 First Lecture on Entropy

Example 1.1 Consider the unilateral shifts $f : \{1, 2\}^\mathbb{N} \to \{1, 2\}^\mathbb{N}$ and $g : \{1, 2, 3\}^\mathbb{N} \to \{1, 2, 3\}^\mathbb{N}$ equipped with the Bernoulli measures $\mu$ and $\nu$, respectively. Show that, for any set $X \subset \{1, 2\}^\mathbb{N}$ satisfying $f^{-1}(X) = X$ and $\mu(X) = 1$, there exists $x \in X$ such that $\#(X \cap f^{-1}(x)) = 2$. Conclude that if $k \neq l$ then $(f, \mu)$ and $(g, \nu)$ cannot be ergodically equivalent.

Proof. Denote $X_i = X \cap [0; i]$ and $p_i = \mu([0; i])$, for $i = 1, 2$. As $\mu$ is a Bernoulli measure, $f(X_1) = f(X_2) = f(X)$ has full $\mu$ measure. Now take $x \in f(X)$. So, $x$ has two pre-images. Similarly, for the system $(g, \nu)$, we can find a set $g(Y_1) \cap g(Y_2) \cap g(Y_3)$ with full measure, such that any point $y$ in it has three pre-images. From here, one can easily show that it is impossible for these two systems to be ergodically equivalent. 

Unfortunately, we do not have such a nice method for bilateral shifts. As it turns out, entropy is a new invariant to distinguish bilateral shifts.

Now we make definitions. Suppose that $(M, \mathcal{B}, \mu)$ is a probability space.

Definition 1.2 Given a finite or countable partition $\mathcal{P}$, the information function associated with this partition is defined as

\[ I_{\mu, \mathcal{P}}(x) = -\sum_{A \in \mathcal{P}} \chi_A(x) \log \mu(A) \quad (1.1) \]

Definition 1.3 We call the entropy of the above partition for the integration of the information function over the whole space:

\[ H_{\mu}(\mathcal{P}) = \int_M I_{\mathcal{P}}(x) d\mu(x) = \sum_{A \in \mathcal{P}} -\mu(A) \log \mu(A) \quad (1.2) \]

In other words, entropy is defined as some kind of average information one get in the whole probability space as one assigns all the points with countable different names in order to distinguish them.

Definition 1.4 More generally, given a sub-$\sigma$-algebra $\mathcal{A}$ of $\mathcal{B}$, we define the conditional information of $\mathcal{P}$ given $\mathcal{A}$ as the function

\[ I_{\mu, \mathcal{P}/\mathcal{A}}(x) = -\sum_{A \in \mathcal{P}} \mu(A/\mathcal{A})(x) \log \mu(A/\mathcal{A})(x) \quad (1.3) \]

where, the conditional probability is defined as $\mu(A/\mathcal{A})(x) := E(\chi_A/\mathcal{A})(x)$. 

1
Moreover, the conditional entropy of \( \mathcal{P} \) given the sub-\( \sigma \)-algebra \( \mathcal{A} \) is defined as

\[
H_\mu(\mathcal{P}/\mathcal{A}) = \int I_{\mathcal{P}/\mathcal{A}}(x)d\mu(x) = -\int \sum_{A \in \mathcal{P}} \mu(A/\mathcal{A}) \log \frac{\mu(A/\mathcal{A})}{\mu(A)} d\mu(x) \tag{1.4}
\]

**Exercise 1.1** Suppose \( \mathcal{P} \) is a countable partition and \( \mathcal{A} \) is a sub-\( \sigma \)-algebra as above, then \( H_\mu(\mathcal{P}/\mathcal{A}) = 0 \) if and only if any element \( P \in \mathcal{P} \) is \( \mathcal{A} \)-measurable.

**Exercise 1.2** Suppose \( \mathcal{A} \) is the sub-\( \sigma \)-algebra generated by the countable partition \( \mathcal{Q} \). Deduce that in this case the conditional entropy has the following form:

\[
H_\mu(\mathcal{P}/\mathcal{Q}) = \sum_{P \in \mathcal{P}} \sum_{Q \in \mathcal{Q}} -\mu(P \cap Q) \log \frac{\mu(P \cap Q)}{\mu(Q)}. \tag{1.5}
\]

Suppose \( \mathcal{P}, \mathcal{Q}, \mathcal{R} \) are three countable partitions. Denote by \( \mathcal{P} \lor \mathcal{Q} \) the countable partition consisting of sets \( A \cap B \) with \( A \in \mathcal{P} \) and \( B \in \mathcal{Q} \). A basic and important identity for both the information function and the entropy are as follows:

**Proposition 1.5**

1. \( I_{\mu, \mathcal{P} \lor \mathcal{Q}/\mathcal{R}}(x) = I_{\mu, \mathcal{P}/\mathcal{R}}(x) + I_{\mu, \mathcal{Q}/\mathcal{P} \lor \mathcal{R}}(x) \)
2. \( H_{\mu}(\mathcal{P} \lor \mathcal{Q}/\mathcal{R}) = H_{\mu}(\mathcal{P}/\mathcal{R}) + H_{\mu}(\mathcal{Q}/\mathcal{P} \lor \mathcal{R}) \)

**Proof.**

**Theorem 1.6 (Jensen’s Inequality)** Let \( \varphi \) be a continuous concave function defined on the unit interval \([0, 1]\), that is, \( \varphi(ax + by) \geq a\varphi(s) + b\varphi(y) \) for any \( a, b, x, y \in [0, 1] \) with \( a + b = 1 \). Suppose \( f : (M, \mathcal{B}, \mu) \rightarrow [0, 1] \), then for any sub-\( \sigma \)-algebra \( \mathcal{A} \) of \( \mathcal{B} \),

\[
\varphi \circ \mathbb{E}(f|\mathcal{A})(x) \geq \mathbb{E}(\varphi \circ f|\mathcal{A})(x)
\]

holds for \( \mu \)-a.e. point \( x \).

**Proof.** Suppose \( f \) is a simple function mapping \( X \) into \([0, 1]\); \( f(x) = \sum_i b_i \chi_{B_i}(x) \) where \( \{B_i\} \) is a partition of \( X \). Then \( \sum_i \mu(B_i/\mathcal{A}) = 1 \) and \( \sum_i \mu(B_i/\mathcal{A})\varphi(b_i) \leq \varphi(\sum_i b_i\mu(B_i/\mathcal{A})) \) by concavity. Then standard approximation method implies the same is true for general measurable function \( f \).

**Exercise 1.3** Apply Jensen’s inequality to prove that, if \( \mathcal{P} \) is a countable partition, and given two sub-\( \sigma \)-algebras \( \mathcal{A}_1 \supset \mathcal{A}_2 \), then \( H_\mu(\mathcal{P}/\mathcal{A}_1) \leq H_\mu(\mathcal{P}/\mathcal{A}_2) \).

**Definition 1.7** Given a partition \( \mathcal{P} \) of \( M \), and \( n \geq 1 \), define \( \mathcal{P}^n = \lor_{i=0}^{n-1} f^{-i}(\mathcal{P}) \). Then we have \( H_\mu(\mathcal{P}^n) \) is a subadditive, and thus, define

\[
h_\mu(f, \mathcal{P}) := \lim_{n \to \infty} \frac{1}{n} H_\mu(\mathcal{P}^n)
\]

Finally, define the entropy of a dynamical system as,

\[
h_\mu(f) := \sup_{\mathcal{P}} h_\mu(f, \mathcal{P})
\]

where the supremum is taken over all the partitions with finite entropy.
Lemma 1.8 If $Q$ is a countable partition with $H_\mu(Q) < \infty$, and if $P_1 \prec P_2 \prec \cdots$ is an increasing sequence of partitions. Denote by $\varphi(x) := \sup_n I_{\mu, Q/P_n}(x)$, then one has the estimate:

$$\int_M \varphi(x) d\mu(x) \leq H_\mu(Q) + 1$$

Proof. Consider the function defined by $F(t) = \mu \{ x \in M : \varphi(x) > t \}$. It follows that

$$F(t) = \mu \left\{ x \in M : \sum_n \chi_A(x) \log \mu(A/P_n)(x) > t \right\}$$

$$= \sum_{A \in Q} \mu(A \cap \left\{ x : \inf_n \mu(A/P_n) < e^{-t} \right\})$$

$$= \sum_{A \in Q} \sum_{n=1}^\infty \mu(A \cap \left\{ x : \mu(A/P_n) < e^{-t}, \mu(A/P_k) \geq e^{-t} \text{ for all } k < n \right\})$$

$$= \sum_{A \in Q} \sum_{n=1}^\infty \int_{\{x : \mu(A/P_n) < e^{-t}, \mu(A/P_k) \geq e^{-t} \text{ for all } k < n\}} \mu(A/P_n)(x) d\mu(x)$$

$$\leq \sum_{A \in Q} \min(m(A), e^{-t})$$

Then, we have that

$$\int_M \varphi(x) d\mu(x) = \int_0^\infty F(t) dt$$

$$\leq \sum_{A \in Q} \int_0^\infty \min(m(A), e^{-t}) dt$$

$$\leq \sum_{A \in Q} (-\mu(A) \log \mu(A) + \mu(A))$$

$$= H_\mu(Q) + 1$$

Lemma 1.9 Suppose that $P_1 \prec P_2 \prec \cdots$ is a sequence of non-decreasing partitions, and the finite partition $Q$ consists of sets in the $\sigma$-algebra generated by $\bigcup_{n=1}^\infty P_n$, then, $\lim_n H_\mu(Q/P_n) = 0$.

Proof. Denote the $\sigma$-algebra generated by the sequence by $\mathcal{A}$. Consider the sequence of functions $I_{\mu, Q/P_n}(x)$. By the martingale convergence theorem, we have that $I_{\mu, Q/P_n}(x) \to I_{\mu, Q/A}(x) = 0$ for $\mu$-a.e point $x$, because $Q$ consists of sets in $\mathcal{A}$. Finally, Note that $H_\mu(Q/P_n) = \int_M I_{\mu, Q/P_n}(x) d\mu(x)$. In view of lemma 1.8, and use the dominated convergence theorem, one gets $H_\mu(Q/P_n) \to H_\mu(Q/A) = 0$.

Theorem 1.10 (Kolmogorov-Sinai) Let $P_1 \prec \cdots \prec P_n \prec \cdots$ be a non-decreasing sequence of partitions with finite entropy such that $\bigcup_{n=1}^\infty P_n$ generates the $\sigma$-algebra of measurable sets. Then

$$h_\mu(f) = \lim_n h_\mu(f, P_n).$$

Proof. We claim that for any two partitions $P$ and $Q$, we have

$$h_\mu(f, Q) \leq h_\mu(f, P_n) + H_\mu(Q/P_n) \text{ for any } n.$$
In fact, for any $n \geq 1$ it holds
\[
H_\mu(Q^n/P^n) = H_\mu(Q^{n-1} \cup f^{-(n-1)}(Q)/P^{n-1} \cup f^{-(n-1)}(P)) \\
\leq H_\mu(Q^{n-1}/P^{n-1}) + H_\mu(f^{-(n-1)}(Q)/f^{-(n-1)}(P)) \\
= H_\mu(Q^{n-1}/P^{n-1}) + H_\mu(Q/P)
\]
which implies that
\[
H_\mu(Q^n/P^n) \leq nH_\mu(Q/P) \quad \text{for any } n \geq 1.
\]
It follows that
\[
H_\mu(Q^n) \leq H_\mu(P^n \cup Q^n) = H_\mu(P^n) + H_\mu(Q^n/P^n) \leq H_\mu(P^n) + nH_\mu(Q/P).
\]
Dividing by $n$ and passing to the limit as $n \to \infty$ we get the conclusion of the lemma.

Passing to the limit as $n \to \infty$ and then taking the supremum over all the finite partitions $Q$ we obtain the conclusion of the theorem.

**Example 1.11** Let $M = \{1, \ldots, d\}^N$ be equipped with a product measure $\mu = \nu^N$. Denote by $p_i = \nu(\{i\})$ for each $i \in \{1, \ldots, d\}$. For any $n \geq 1$, let $P_n$ be the partition of $M$ into cylinders $[0; a_0, \ldots, a_{n-1}]$ of length $n$. The entropy of $P_n$ is $H_\mu(P_n) = -n \sum d p_i \log p_i$, then
\[
\lim_{n \to \infty} h_\mu(f) = h_\mu(f, P_n)
\]
and,
\[
\lim_{n \to \infty} \frac{1}{n} H_\mu(P_n) = H_\mu(P^n) = -\sum d p_i \log p_i
\]

## 2 Examples; Local Entropy

**Corollary 2.1** Let $P$ be a partition with finite entropy such that the union of its iterates $P^n = \vee_{j=0}^{n-1} f^{-j}(P)$, $n \geq 1$, generates the $\sigma$-algebra of measurable sets. Then $h_\mu(f) = h_\mu(f, P)$.

**Proof.** Apply Kolmogorov-Sinai theorem to the sequence $P^n$. Recall that $h_\mu(f, P^n) = h_\mu(f, P)$ for any $n$.

**Definition 2.2** In this case, we say the partition is unilateral generating.

(Example 1.11, again) Consider the Bernoulli shift $f$ on $M = \{1, \ldots, d\}^N$ with a product measure $\mu = \nu^N$. Denote by $p_i = \nu(\{i\})$ for each $i \in \{1, \ldots, d\}$. Consider the partition $P$ of cylinders $[0; i]$. Then
\[
\lim_{n \to \infty} h_\mu(f) = h_\mu(f, P)
\]
and,
\[
\lim_{n \to \infty} \frac{1}{n} H_\mu(P^n) = -\sum d p_i \log p_i
\]

**Example 2.3** Let $\Sigma = \{1, \ldots, d\}^N$ and let $\sigma : \Sigma \to \Sigma$ be the shift. Let $\mu$ be the Markov measure associated with a stochastic matrix $P = (P_{i,j})_{i,j}$ and a probability vector $p = (p_i)_i$. The entropy of this system is as follows:
\[
h_\mu(\sigma) = \sum_{a=1}^d p_a \sum_{b=1}^d -P_{a,b} \log P_{a,b}
\]
Proof. Consider the partition $\mathcal{P}$ of $M$ into cylinders $[0; a]$, $a = 1, \ldots, d$. For each $n$, the iterate $\mathcal{P}^n$ is the partition into cylinders $[0; a_1, \ldots, a_n]$ of length $n$. Recall that $\mu([0; a_1, \ldots, a_n]) = p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n}$, and we see that

$$H_\mu(\mathcal{P}^n) = \sum_{a_1, \ldots, a_n} -p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \log (p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n})$$

$$= \sum_{a_1} -p_{a_1} \log p_{a_1} \sum_{a_2, \ldots, a_n} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n}$$

$$+ \sum_{j=1}^{n-1} \sum_{a_j, a_{j+1}} -\log P_{a_j, a_{j+1}} \sum_{a_1} p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n},$$

where the last sum is over all the values of $a_1, \ldots, a_{j-1}, a_{j+2}, \ldots, a_n$. On one hand,

$$\sum_{a_2, \ldots, a_n} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} = \sum_{a_n} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} = 1$$

since $P^n$ is a stochastic matrix. Similarly, recall that $\mathcal{P}^* p = p$,

$$\sum p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} = \sum_{a_1, \ldots, a_n} p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} = \sum_{a_1, \ldots, a_n} p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n}$$

$$= \sum_{a_1} p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} = p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n}.$$

For the last equality, recall that $\mathcal{P}^* p = p = p$. Replacing this in (2.1), we obtain that

$$H_\mu(\mathcal{P}^n) = \sum_{a_1} -p_{a_1} \log p_{a_1} + \sum_{j=1}^{n-1} \sum_{a_j, a_{j+1}} -p_{a_j} P_{a_j, a_{j+1}} \log P_{a_j, a_{j+1}}$$

$$= \sum_{a_1} -p_{a_1} \log p_{a_1} + (n-1) \sum_{a, b} -p_{a} P_{a, b} \log P_{a, b}.$$

Then $h_\mu(f, \mathcal{P}) = \sum_{a, b} -p_{a} P_{a, b} \log P_{a, b}$. As the family of all the cylinders $[0; a_1, \ldots, a_n]$ generates the $\sigma$-algebra of $M = \{1, \ldots, d\}^\mathbb{N}$, it follows from Corollary 2.1 that $h_\mu(f) = h_\mu(f, \mathcal{P})$. This completes the computation.

Corollary 2.4 Suppose that the system $(f, \mu)$ is invertible. Let $\mathcal{P}$ be a partition with finite entropy such that the union of the iterates $\mathcal{P}^{\pm n} = \bigvee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$, generates the $\sigma$-algebra of measurable sets. Then $h_\mu(f) = h_\mu(f, \mathcal{P}).$

Proof. Apply Kolmogorov-Sinai theorem to the sequence $\mathcal{P}^{\pm n}$. Recall that $h_\mu(f, \mathcal{P}^{\pm n}) = h_\mu(f, \mathcal{P})$ for any $n$.

Definition 2.5 In this case we say that the partition $\mathcal{P}$ is bilateral generating.

Note the for bilateral Bernoulli shift, we can use this notation to compute its entropy.

Corollary 2.6 Suppose now the space $M$ is a metric space. Let $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ be a non-decreasing sequence of partitions with finite entropy such that $\text{diam} \mathcal{P}_n(x) \to 0$ for $\mu$-almost every point $x \in M$. Then

$$h_\mu(f) = \lim_{n} h_\mu(f, \mathcal{P}_n).$$
Proof. Let $U$ be any open set of $M$. The hypothesis guarantees that for any $x$ there exists $n(x)$ such that the set $P_x = \mathcal{P}_{n(x)}(x)$ is contained in $U$. It is clear that $P_x$ belongs to the algebra $\mathcal{A}$ generated by $\cup_x \mathcal{P}_n$. Observe also that this algebra is countable, since it is formed by the finite unions of the elements of the partitions of $\mathcal{P}_n$. In particular, the set of the values taken by $P_x$ is countable. It follows that $U = \cup_{x \in U} P_x$ is also in this algebra $\mathcal{A}$. This proves that the $\sigma$-algebra generated by $\mathcal{A}$ contains all the open sets and, therefore, contains all the Borel sets. Now, the conclusion follows from a direct application of Kolmogorov-Sinai theorem.

Example 2.7 Let $f : S^1 \to S^1$ be a homeomorphism and let $\mu$ be any $f$-invariant probability measure. Given a finite partition $\mathcal{P}$ of $S^1$ into, say, $m$ subintervals, denote by $x_1, \ldots, x_m$ their extreme points. For any $j \geq 1$, the partition $f^{-j}(\mathcal{P})$ is formed by subintervals of $S^1$ determined by the points $f^{-j}(x_i)$. This implies that, for any $n \geq 1$, the set of extreme points of elements of $\mathcal{P}^n$ is

$$\{f^{-j}(x_i) : j = 0, \ldots, n-1 \text{ and } i = 1, \ldots, m\}.$$ 

In particular, $\#\mathcal{P}^n \leq mn$. Then,

$$h_\mu(f, \mathcal{P}) = \lim_{n \to \infty} -\frac{1}{n} H_\mu(\mathcal{P}^n) \leq \lim_{n \to \infty} -\frac{1}{n} \log \#\mathcal{P}^n = \lim_{n \to \infty} -\frac{1}{n} \log mn = 0.$$ 

Now consider any increasing sequence of finite partitions with diameter going to zero and apply Corollary 2.6, and it follows that $h_\mu(f) = 0$:

Corollary 2.8 Let $\mathcal{P}$ be a partition with finite entropy such that, for $\mu$-almost every point $x \in M$, we have that $\text{diam} \mathcal{P}^n(x) \to 0$. Then $h_\mu(f) = h_\mu(f, \mathcal{P})$.

Proof. Just apply Corollary 2.6 to the sequence $\mathcal{P}^n$, keeping in mind that $h_\mu(f, \mathcal{P}^n) = h_\mu(f, \mathcal{P})$ for any $n$.

Theorem 2.9 (Shannon-McMillan-Breiman) Given any partition $\mathcal{P}$ with finite entropy, the limit

$$h_\mu(f, \mathcal{P}, x) = \lim_{n \to \infty} -\frac{1}{n} \log \mu(\mathcal{P}^n(x))$$

exists for $\mu$-almost every point. \hspace{1cm} (2.2)

The function $x \mapsto h_\mu(f, \mathcal{P}, x)$ is $\mu$-integrable, and the identity also holds in $L^1(\mu)$. Moreover,

$$\int h_\mu(f, \mathcal{P}, x) d\mu(x) = h_\mu(f, \mathcal{P}).$$

If $(f, \mu)$ is ergodic then $h_\mu(f, \mathcal{P}, x) = h_\mu(f, \mathcal{P})$ for $\mu$-almost every point.

Proof. Denote $\varphi(x) = I_{\mu, \mathcal{P}/\cup_{n=1}^{\infty} \mathcal{P}^n}(x)$ and observe that $h_\mu(f, \mathcal{P}) = \int_X \varphi(x) d\mu(x)$. We claim that the limit at the left-hand side of the equation (2.2) is $\mathbb{E}(\varphi/\mathcal{C})$, where $\mathcal{C}$ is the sub-$\sigma$-algebra of $f$-invariant sets. Note also that $-\frac{1}{n} \log \mu(\mathcal{P}^n(x)) = \frac{1}{n} I_{\mu, \mathcal{P}^n}(x)$. So we need to prove the following:

$$\lim_{n \to \infty} -\frac{1}{n} I_{\mu, \mathcal{P}^n}(x) = \mathbb{E}(\varphi/\mathcal{C})(x)$$

both a.e. and in $L^1$. From the claim, other conclusions of the theorem are straightforward.

We start with the following estimate:

$$I_{\mu, \mathcal{P}^n}(x) = I_{\mu, \mathcal{P}^n/f^{-1}(\mathcal{P}^{n-1})}(x) + I_{\mu, f^{-1}(\mathcal{P}^{n-1})}(x)$$

$$= I_{\mu, \mathcal{P}/f^{-1}(\mathcal{P}^{n-1})}(x) + I_{\mu, \mathcal{P}^{n-1}} \circ f(x)$$

$$= I_{\mu, \mathcal{P}/f^{-1}(\mathcal{P}^{n-1})}(x) + I_{\mu, \mathcal{P}/f^{-1}(\mathcal{P}^{n-2})} \circ f(x) + \cdots + I_{\mu, \mathcal{P}/f^{-1}(\mathcal{P})} \circ f^{n-2}(x) + I_{\mu, \mathcal{P}} \circ f^{n-1}(x)$$


6
Since we have $\frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \to \mathbb{E}(\varphi/C)$ a.e. and in $L^1$, it suffices to show that
\[
\frac{1}{n} \sum_{j=0}^{n-1} |I_{\mu, P_{f^{-1}(P^{n-1-j})}} \circ f^j(x) - \varphi \circ f^j(x)| \to 0
\]
both a.e. and in $L^1$ norm.

For the $L^1$ convergence, note that
\[
\frac{1}{n} \sum_{j=0}^{n-1} \int |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| \circ f^j(x) d\mu(x) \\
= \frac{1}{n} \sum_{j=0}^{n-1} \int |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| d\mu(x) \\
= \frac{1}{n} \sum_{j=0}^{n-1} \int |I_{\mu, P_{f^{-1}(P^{n-1-j})}}| d\mu(x)
\]

By previous arguments to show Kolmogorov-Sinai theorem, we know $I_{\mu, P_{f^{-1}(P^{j})}} \rightarrow \varphi$ a.e. and is bounded by an integrable function. So $\int |I_{\mu, P_{f^{-1}(P^{j})}}(x) - \varphi(x)| d\mu(x) \rightarrow 0$, which implies the conclusion.

Now we define the function $G_N = \sup_{j \geq N} |I_{\mu, P_{f^{-1}(P^{j})}}(x) - \varphi(x)|$. Again, $G_N$ is bounded by a $L^1$ integrable function, and also, $G_N \to 0$ a.e.. So we have $\int_M G_N(x) d\mu(x) \to 0$.

Now we estimate the sequence, pointwisely, according to some fixed large integer $N$, as follows
\[
\frac{1}{n} \sum_{j=0}^{n-1} |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| \circ f^j(x)
= \frac{1}{n} \sum_{j=0}^{n-N-1} |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| \circ f^j(x) + \frac{1}{n} \sum_{j=n-N}^{n-1} |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| \circ f^j(x)
\leq \frac{1}{n} \sum_{j=0}^{n-N-1} G_N \circ f^j(x) + \frac{1}{n} \sum_{j=n-N}^{n-1} G_0 \circ f^j(x)
\]
So we have almost everywhere,
\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| \circ f^j(x) \leq \mathbb{E}(G_N/C)
\]

Finally, since $\mathbb{E}(G_N/C) \to 0$ a.e., one concludes that $\limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} |I_{\mu, P_{f^{-1}(P^{n-1-j})}} - \varphi| \circ f^j(x) = 0$ a.e.

**Example 1.11, once again** Consider the Bernoulli shift on $\Sigma = \{1, \ldots, d\}^N$. Let $\mu$ be the Bernoulli measure associated with the probability vector $(p_1, \ldots, p_d)$. Define $\eta_i(x, n) = \#\{0 \leq j \leq n - 1 : x_j = i\}$. Let $P$ be the (generating) partition of $\Sigma$ into cylinders $[0; i]$. Note that
\[
\mu(P^n(x)) = \prod_{i=1}^d p_i^{\eta_i(x, n)}.
\]
By ergodicity, $\eta_i(x, n)/n \to p_i$ at $\mu$-almost every point $x$. Thus, $-(1/n) \log \mu(P^n(x))$ converges to $\sum_{i=1}^d -p_i \log p_i$ at $\mu$-almost every point.
Definition 2.10 Suppose that $f : M \to M$ is a continuous map on a compact metric space. Given $x \in M$, $n \geq 1$ and $\varepsilon > 0$, we call a dynamical ball of length $n$ and radius $\varepsilon$ at $x$ for the set:

$$B(x, n, \varepsilon) = \{ y \in M : d(f^j(x), f^j(y)) < \varepsilon \text{ for any } j = 0, 1, \ldots, n-1 \}.$$ 

In other words, $B(n, x, \varepsilon) = \cap_{j=0}^{n-1} f^{-j}(B(f^j(x), \varepsilon))$. Define:

$$h^+_\mu(f, x, \varepsilon) = \limsup_n -\frac{1}{n} \log \mu(B(x, n, \varepsilon))$$

$$h^-_\mu(f, x, \varepsilon) = \liminf_n -\frac{1}{n} \log \mu(B(x, n, \varepsilon)).$$

Theorem 2.11 (Brin-Katok) Suppose that $f : M \to M$ is a continuous map on a compact metric space. Let $\mu$ be an $f$-invariant measure, the limits

$$\lim_{\varepsilon \to 0} h^+_\mu(f, x, \varepsilon) \quad \text{and} \quad \lim_{\varepsilon \to 0} h^-_\mu(f, x, \varepsilon)$$

exist and are equal for $\mu$-almost every point. Denoting by $h_\mu(f, x)$ for their common value, the function $h_\mu(f, \cdot)$ is integrable and we have

$$h_\mu(f) = \int h_\mu(f, x) d\mu(x).$$

Example 2.12 Let $f_A : \mathbb{T}^d \to \mathbb{T}^d$ be the endomorphism on the torus $\mathbb{T}^d$ induced by some invertible matrix $A$ with integer coefficients. Suppose further that the Lebesgue measure $\mu$ on $\mathbb{T}^d$ is preserved by the endomorphism $f_A$. Then

$$h_\mu(f_A) = \sum_{i=1}^d \log^+ |\lambda_i|.$$ 

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $A$, counted with multiplicity.

We only show the case when the matrix $A$ is diagonalizable, with real eigenvalues. Let $e_1, \ldots, e_d$ be an orthonormal base of $\mathbb{R}^d$ such that $Ae_i = \lambda_i e_i$ for any $i$. Let $u$ be the number of eigenvalues of $A$ with absolute value strictly greater than 1. We can suppose that the eigenvalues are indexed in such a way that $|\lambda_i| > 1$ if and only if $i \leq u$. Given $x \in \mathbb{T}^d$, any point $y$ in a neighborhood of $x$ can be written in the form

$$y = x + \sum_{i=1}^d t_i e_i$$

with $t_1, \ldots, t_d$ close to zero. Given $\varepsilon > 0$, denote by $D(x, \varepsilon)$ the set of points $y$ in this form with $|t_i| < \varepsilon$ for any $i = 1, \ldots, d$. Moreover, for any $n \geq 1$, consider

$$D(x, n, \varepsilon) = \{ y \in \mathbb{T}^d : f_A^j(y) \in D(f_A^j(x), \varepsilon) \text{ for any } j = 0, \ldots, n-1 \}.$$ 

Observe that $f_A^j(y) = f_A^j(x) + \sum_{i=1}^d t_i \lambda_i^j e_i$ for any $n \geq 1$. Therefore,

$$D(x, n, \varepsilon) = \{ x + \sum_{i=1}^d t_i e_i : |\lambda_i^n t_i| < \varepsilon \text{ for } i \leq u \text{ and } |t_i| < \varepsilon \text{ for } i > u \}.$$ 

So $\mu(D(x, n, \varepsilon)) = \varepsilon^d \prod_{i=1}^u |\lambda_i|^{-n}$ for any $x \in \mathbb{T}^d$, $n \geq 1$ and $\varepsilon > 0$. We can also choose $C > 1$, such that

$$B(x, C^{-1}\varepsilon) \subset D(x, \varepsilon) \subset B(x, C\varepsilon)$$
for any \( x \in \mathbb{T}^d \) and any \( \varepsilon > 0 \) sufficiently small. Then, \( D(x, n, \varepsilon/C) \subset B(x, n, \varepsilon) \subset D(x, n, C\varepsilon) \) for any \( n \geq 1 \). Then we obtain that:

\[
C^{-d} \varepsilon^d \prod_{i=1}^{n} |\lambda_i|^{-n} \leq \mu(B(x, n, \varepsilon)) \leq C^d \varepsilon^d \prod_{i=1}^{n} |\lambda_i|^{-n}
\]

for any \( x \in \mathbb{T}^d \), \( n \geq 1 \) and \( \varepsilon > 0 \). Then,

\[
h_{\mu}^+(f, \varepsilon, x) = h_{\mu}^-(f, \varepsilon, x) = \lim_{n} \frac{1}{n} \log \mu(B(x, n, \varepsilon)) = \sum_{i=1}^{n} \log |\lambda_i|
\]

for any \( x \in \mathbb{T} \) and any small \( \varepsilon > 0 \). So, \( h_{\mu}(f, x) = \sum_{i=1}^{n} |\lambda_i| \) and, therefore, by the Brin-Katok theorem (Theorem 2.11), it follows

\[
h_{\mu}(f) = \int h_{\mu}(f, x)d\mu(x) = \sum_{i=1}^{n} \log |\lambda_i|
\]

### 3 Jacobian and Rokhlin’s Formula

**Example 3.1** Let \( f : U \to f(U) \) be a diffeomorphism of an open subset \( U \) of \( \mathbb{R}^d \), and let \( m \) denote the usual Lebesgue measure. For any integrable function \( \varphi(y) \) defined on \( f(U) \), by the change of variables formula, we have the following identity:

\[
\int_{f(U)} \varphi(y)m(y) = \int_{U} \varphi \circ f(x)|\det(Df(x))|d\mu(x)
\]

Let \( f : M \to M \) be a measurable map on a metric space \( M \) and suppose that \( f \) is locally invertible: any point \( x \in M \) admits a neighborhood \( U_x \) such that the restriction \( f \mid U_x \) is a bijection onto its image, which is a measurable set, and the inverse of this bijection is also measurable. We call injectivity domain of the map for any measurable subset of such a neighborhood. Note that, by definition, if \( A \) is an injectivity domain then \( f(A) \) is a measurable set.

**Definition 3.2** Let \( \eta \) be a probability measure on \( M \). A measurable function \( \xi : M \to [0, \infty) \) is a Jacobian of \( f \) with respect to \( \eta \) if the restriction of \( \xi \) to any injectivity domain \( A \) is integrable with respect to \( \eta \) and satisfies

\[
\eta(f(A)) = \int_{A} \xi d\eta.
\]

In this case, we denote by \( J_{\eta}f \) for this function \( \xi \).

**Exercise 3.1** Let \( f : M \to M \) be a locally invertible map and \( \eta \) be a Borel probability measure on \( M \) which is not singular with respect to \( f \). Show that it holds the following change of variable formula:

(a) \( \int_{f(A)} \varphi d\eta = \int_{A} (\varphi \circ f)J_{\eta}f d\eta \) for any invertibility domain \( A \subset M \) and any measurable function \( \varphi : f(A) \to \mathbb{R} \) such that the integrals is defined (which can take \( \pm \infty \)).

(b) \( \int_{A} \psi d\eta = \int_{f(A)} (\psi/J_{\eta}f) \circ (f \mid A)^{-1} d\eta \) for any measurable function \( \psi : A \to \mathbb{R} \) such that the integrals are defined (which can take \( \pm \infty \)).
Definition 3.3 We say that a measure \( \eta \) is non-singular with respect to the map \( f \) if the image of any subset of an injectivity domain with zero measure also has zero measure: if \( \eta(A) = 0 \) then \( \eta(f(A)) = 0 \). For example, if \( f : U \to f(U) \) is a diffeomorphism on an open set \( U \subset \mathbb{R}^d \) and \( \eta \) is the Lebesgue measure then \( \eta \) is non-singular. It is also easy to verify that any invariant probability measure is non-singular.

Propostion 3.4 Let \( M \) be a separable metric space and let \( f : M \to M \) be a locally invertible map. Then, given any Borel measure \( \eta \) on \( M \) which is non-singular with respect to \( f \), there exists some Jacobian of \( f \) with respect to \( \eta \). Moreover, the Jacobian is essentially unique: any two Jacobians coincide for \( \eta \)-almost every point.

**Proof.** Take some countable partition \( \mathcal{P} \) of the space \( M \), each part \( A \in \mathcal{P} \) is an injectivity domain of \( f \). For each \( P \in \mathcal{P} \), denote by \( \eta_P \) the measure defined on \( P \) by \( \eta_P(A) = \eta(f(A)) \). The hypothesis that \( \eta \) is non-singular implies that each \( \eta_P \) is absolutely continuous with respect to \( \eta \) restricted to \( P \):

\[
\eta(A) = 0 \implies \eta(f(A)) = \eta_P(A) = 0
\]

for any measurable set \( A \subset P \). Let \( \xi_P(x) = d\eta_P/d\eta(P) \) be the Radon-Nykodym derivative. Then \( \xi_P(x) \) is a function defined on \( P \), integrable with respect to \( \eta \) and satisfies

\[
\eta(f(A)) = \eta(P(A)) = \int_A \xi_P \, d\eta
\]

for any measurable set \( A \subset P \). Consider the function \( \xi : M \to [0, \infty) \) whose restriction to each \( P \in \mathcal{P} \) is given by \( \xi_P(x) \). Then, for any subset in an injectivity domain, we have

\[
\eta(f(A)) = \int_A \xi(x) \, d\eta(x)
\]

This proves that \( \xi \) is a Jacobian of \( f \) with respect to \( \eta \).

Now suppose that \( \xi \) and \( \zeta \) are two Jacobians of \( f \) with respect to \( \eta \). Then for any small set \( B \subset M \), one has that \( \eta(f(B)) = \int_B \xi(x) \, d\eta(x) = \int_B \zeta(x) \, d\eta(x) \). This implies that \( \xi(x) = \zeta(x) \) almost everywhere.

Theorem 3.5 (the Rohklin formula) Consider the measure-preserving system \((M, f, \mu)\) where \( M \) is a separable metric space and \( f \) is locally invertible. Suppose that there exists some finite or countable partition \( \mathcal{P} \) such that any element \( P \in \mathcal{P} \) is an injectivity domain of \( f \) and \( \text{diam} P^n(x) \to 0 \) for \( \mu \)-almost every \( x \). Then \( h_\mu(f) = \int \log J_\mu f(x) \, d\mu(x) \).

**Proof.** Given any bounded measurable function \( \psi : M \to \mathbb{R} \), we make the following computation:

\[
\int_M \psi(y) \, d\mu(y) = \int_M \psi \circ f(x) \, d\mu(x) = \sum_{B \in \mathcal{P}} \int_B \psi \circ f(x) \, d\mu(x) = \sum_{B \in \mathcal{P}} \int_{f(B)} \psi \circ f \circ J_\mu f^{-1}(y) \, d\mu(y) = \int_M \sum_{z : \psi(z) = \psi(f(y))} \frac{\psi}{J_\mu f}(z) \, d\mu(y)
\]
It is not hard to deduce that \( \mu_k \) for any compute the entropy: In particular, when we take \( \psi(x) = \chi_A \) for some \( A \in \mathcal{P} \), restricted to which set the map is invertible, we have that \( \mathbb{E}(\chi_A/ \cap_{n=1}^{\infty} f^{-n}(\mathcal{P})) = \chi_A(y) \). Now we know that, by Kolmogorov-Sinai theorem,

\[
h(\mu) = h(\mu, \mathcal{P}) = \lim_{n \to \infty} H(\mu, \cap_{j=1}^{n} f^{-j}(\mathcal{P})) = H(\mu, \cap_{n=1}^{\infty} f^{-n}(\mathcal{P}))
\]

Finally, we compute that

\[
h(\mu) = H(\mu, \cap_{n=1}^{\infty} f^{-n}(\mathcal{P}))
\]

Example 3.6 (Bernoulli shift, the 4th time) Let \( \sigma : \Sigma \to \Sigma \) be the shift on \( \Sigma = \{1, 2, \ldots, d\}^\mathbb{N} \) and let \( \mu \) be the Bernoulli measure associated with the probability vector \( p = (p_1, \ldots, p_d) \). Consider \( \Sigma \) equipped with a metric which is compatible with the product topology. Consider the partition \( \mathcal{P} \) into cylinders \([0; i]\). The restriction of \( \sigma \) on each cylinder \([0; i]\) is an invertible map. Moreover, given any cylinder \([0; a_1, \ldots, a_n] \subset [0; a] \),

\[
\mu(\sigma([0; a_1, \ldots, a_n])) = p_{a_1} \cdots p_{a_n} = \frac{1}{p_a} \mu([0; a_1, \ldots, a_n]).
\]

It is not hard to deduce that \( \mu(\sigma(A)) = \frac{1}{p_a} \mu(A) \) for any measurable subset \( A \subset [0; a] \). Therefore, the function \( J_\mu f(x) = \frac{1}{p_a} \) is a Jacobian of \( \sigma \) with respect to \( \mu \). So, we can apply Rokhlin’s formula to compute the entropy:

\[
h(\mu) = \int \log J_\mu f(x) d\mu(x) = -\sum_{i=1}^{d} p_i \log p_i
\]

Example 3.7 (The Gauss map) The entropy of the Gauss map \( G(x) = \frac{1}{x} - \frac{1}{2} \) with respect to the invariant measure \( d\mu = \frac{1}{\log 2} \frac{dx}{1+2x} \) is \( h(\mu, G) = \int_M \log |G'| d\mu \).

Example 3.8 Let \( f : [0, 1] \to [0, 1] \) be the \( \times 2 \) mod 1 map. It is conjugated to the Bernoulli shift with probability vector \( \left( \frac{1}{2}, \frac{1}{2} \right) \). So the entropy of \( f \) with respect to Lebesgue measure \( m \) is \( h_m(f) = \log 2 \). For any \( k \geq 1 \), denote by \( F_k \) the set of fixed points of the iterate \( f^k \). Observe that \( F_k \) is an invariant
set consisting of \( \#F_k = 2^k - 1 \) points, and that these sets split the interval \([0,1]\) into equi-length subintervals. Note that if we define

\[
\mu_k = \frac{1}{2^k - 1} \sum_{x \in F_k} \delta_x.
\]

The above observations imply that each \( \mu_k \) is an invariant probability measure and that the sequence \((\mu_k)_k\) converges to Lebesgue measure \( m \) in weak* topology. Note that each \( \mu_k \) is supported on a finite set, so one easily computes that \( h_{\mu_k}(f) = 0 \) for any \( k \). Therefore, the entropy does not vary continuously on the invariant measure.

Now we consider the map \( \mu \mapsto h_{\mu}(f) \), for a fixed map \( f \), where \( \mu \) are taken in the space of \( f \)-invariant probability measures. The above example shows that this function could be not continuous. But we have the following theorem:

**Theorem 3.9** Suppose that there exists a finite partition \( \mathcal{P} \) such that \( \mu(\partial \mathcal{P}) = 0 \) and \( \cup_n \mathcal{P}^n \) generates the \( \sigma \)-algebra of measurable subsets of \( M \) up to a set with zero measure. Then the function \( \mu \mapsto h_{\mu}(f) \) is upper semi-continuous at the point \( \mu \).

**Proof.** Since any part of \( \mathcal{P} \) is a continuity set for \( \mu \), that is, its boundary has zero measure. Then the function \( \nu \mapsto \nu(\mathcal{P}) \) is continuous at point \( \mu \), for any \( \mathcal{P} \in \mathcal{P} \). Consequently, the function

\[
\nu \mapsto H_{\nu}(\mathcal{P}) = \sum_{\mathcal{P} \in \mathcal{P}} -\nu(\mathcal{P}) \log \nu(\mathcal{P}),
\]

is also continuous in \( \mu \). Similarly, the invariance of the measure implies that \( \mu(\partial \mathcal{P}^n) = 0 \) for any \( n \geq 1 \), and thus, the function \( \nu \mapsto H_{\nu}(\mathcal{P}^n) \) is continuous for any \( n \).

Then, the function \( h_{\nu}(f,\mathcal{P}) = \lim \frac{1}{n} H_{\nu}(\mathcal{P}^n) = \inf \frac{1}{n} H_{\nu}(\mathcal{P}^n) \) is also upper semi-continuous at the point \( \mu \). Then for any \( \varepsilon > 0 \), there exists an open neighborhood \( V \) of \( \mu \) such that for any \( \nu \in V \), one has \( h_{\nu}(f,\mathcal{P}) \leq h_{\mu}(f,\mathcal{P}) + \varepsilon \), which implies that \( h_{\nu}(f) \leq h_{\mu}(f) + \varepsilon \) for all \( \nu \in V \). \[ \square \]

A continuous map \( f : M \to M \) on a metric space is expansive if there exists \( \varepsilon_0 \) (called the expansivity constant) such that, given \( x, y \in M \) with \( x \neq y \) there exists some \( n \in \mathbb{N} \) such that \( d(f^n(x),f^n(y)) \geq \varepsilon_0 \).

**Exercise 3.2** Let \( f : M \to M \) be an \( \varepsilon_0 \)-expansive map on a compact metric space. Then there exists an \( f \)-invariant probability measure \( \mu \) whose entropy \( h_{\mu}(f) \) is the maximum among all the \( f \)-invariant probability measures.