MAGIC: Ergodic Theory Lecture 7 - Entropy

Charles Walkden

March 6, 2013

Introduction

A central problem in mathematics is the *isomorphism problem*: when are two objects in the same class “the same”?

Introduction

A central problem in mathematics is the *isomorphism problem*: when are two objects in the same class "the same"?
Two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \phi \downarrow & & \downarrow \phi \\ Y & \xrightarrow[S]{} & Y \end{array}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$ (i.e. $\mu(\phi^{-1} B) = m(B)\ \forall B \in \mathcal{A}$).

Introduction

A central problem in mathematics is the *isomorphism problem*: when are two objects in the same class "the same"?
Two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$ (i.e. $\mu(\phi^{-1}B) = m(B)\ \forall B \in \mathcal{A}$). It is natural to look for invariants.

Introduction

A central problem in mathematics is the *isomorphism problem*: when are two objects in the same class "the same"?
Two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \phi \downarrow & & \downarrow \phi \\ Y & \xrightarrow[S]{} & Y \end{array}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$ (i.e. $\mu(\phi^{-1}B) = m(B)\ \forall B \in \mathcal{A}$). It is natural to look for invariants. To each mpt T we will associate a number - its *entropy* $h(T)$.

Introduction

A central problem in mathematics is the *isomorphism problem*: when are two objects in the same class "the same"?
Two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$ (i.e. $\mu(\phi^{-1}B) = m(B)\ \forall B \in \mathcal{A}$). It is natural to look for invariants. To each mpt T we will associate a number - its *entropy* $h(T)$.
If S, T are isomorphic then $h(S) = h(T)$. (Conversely, if $h(S) \neq h(T)$ then S, T cannot be isomorphic.)

Introduction

A central problem in mathematics is the *isomorphism problem*: when are two objects in the same class "the same"?
Two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\phi} \\ Y & \xrightarrow[S]{} & Y \end{array}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$ (i.e. $\mu(\phi^{-1}B) = m(B)\ \forall B \in \mathcal{A}$). It is natural to look for invariants. To each mpt T we will associate a number - its *entropy* $h(T)$.
If S, T are isomorphic then $h(S) = h(T)$. (Conversely, if $h(S) \neq h(T)$ then S, T cannot be isomorphic.)
Throughout, $\log = \log_2$.

Information and entropy of a partition

Information and entropy of a partition

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

Information and entropy of a partition

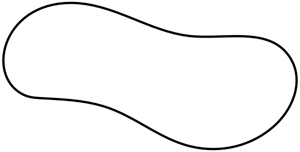
Let $(X, \mathcal{B}, \mu)$ be a probability space.

Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.

Information and entropy of a partition

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

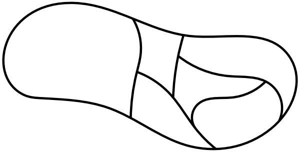
Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.



Information and entropy of a partition

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

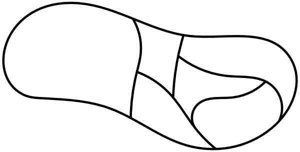
Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.



Information and entropy of a partition

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

Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.

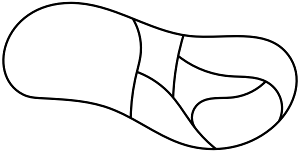


If we know that $x \in A_j$ then we have received some information.

Information and entropy of a partition

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

Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.



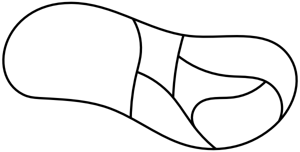
If we know that $x \in A_j$ then we have received some information.

If A_j is 'big' then we have received a 'small' amount of information.

Information and entropy of a partition

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

Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.



If we know that $x \in A_j$ then we have received some information.

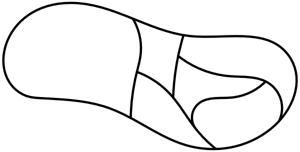
If A_j is 'big' then we have received a 'small' amount of information.

If A_j is 'small' then we have received a 'large' amount of information.

Information and entropy of a partition

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

Suppose we are trying to locate a point $x \in X$ using a partition $\alpha = \{A_j\}$.



If we know that $x \in A_j$ then we have received some information.

If A_j is 'big' then we have received a 'small' amount of information.

If A_j is 'small' then we have received a 'large' amount of information.

This motivates defining the 'information function' as

$$I(\alpha)(x) = \sum_{A \in \alpha} \chi_A(x) \phi(\mu(A))$$

for an appropriate choice of function ϕ.

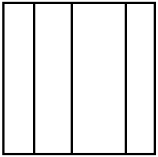
Suppose α and β are two partitions.

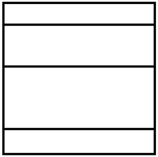
Suppose α and β are two partitions.
The *join* of α and β is the partition

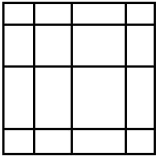
$$\alpha \vee \beta = \{A \cap B \mid A \in \alpha, B \in \beta\}.$$

Suppose α and β are two partitions.
The *join* of α and β is the partition

$$\alpha \vee \beta = \{A \cap B \mid A \in \alpha, B \in \beta\}.$$

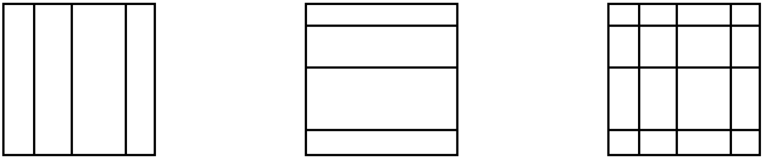

α


β


$\alpha \vee \beta$

Suppose α and β are two partitions.
The *join* of α and β is the partition

$$\alpha \vee \beta = \{A \cap B \mid A \in \alpha, B \in \beta\}.$$



α β $\alpha \vee \beta$

α,β are *independent* if $\mu(A \cap B) = \mu(A)\mu(B)\ \forall A \in \alpha,\ B \in \beta$.

It is natural to assume that if α, β are independent, then:

$$\left|\begin{array}{c}\text{Information obtained} \\ \text{by knowing which} \\ \text{element of } \alpha \vee \beta \text{ we} \\ \text{are in}\end{array}\right| = \left|\begin{array}{c}\text{Information} \\ \text{obtained from } \alpha\end{array}\right| + \left|\begin{array}{c}\text{Information} \\ \text{obtained from } \beta\end{array}\right|$$

It is natural to assume that if α, β are independent, then:

$$\left| \begin{array}{c} \text{Information obtained} \\ \text{by knowing which} \\ \text{element of } \alpha \vee \beta \text{ we} \\ \text{are in} \end{array} \right| = \left| \begin{array}{c} \text{Information} \\ \text{obtained from } \alpha \end{array} \right| + \left| \begin{array}{c} \text{Information} \\ \text{obtained from } \beta \end{array} \right|$$

$$\phi(\mu(A \cap B)) = \phi(\mu(A)\mu(B)) \quad = \quad \phi(\mu(A)) \quad + \quad \phi(\mu(B))$$

It is natural to assume that if α, β are independent, then:

$$\left|\begin{array}{c}\text{Information obtained}\\ \text{by knowing which}\\ \text{element of } \alpha \vee \beta \text{ we}\\ \text{are in}\end{array}\right| = \left|\begin{array}{c}\text{Information}\\ \text{obtained from } \alpha\end{array}\right| + \left|\begin{array}{c}\text{Information}\\ \text{obtained from } \beta\end{array}\right|$$

$$\phi(\mu(A \cap B)) = \phi(\mu(A)\mu(B)) \quad = \quad \phi(\mu(A)) \quad + \quad \phi(\mu(B))$$

This indicates we should take $\phi(t) = -\log t$.

It is natural to assume that if α, β are independent, then:

$$\left|\begin{array}{c}\text{Information obtained}\\ \text{by knowing which}\\ \text{element of } \alpha \vee \beta \text{ we}\\ \text{are in}\end{array}\right| = \left|\begin{array}{c}\text{Information}\\ \text{obtained from } \alpha\end{array}\right| + \left|\begin{array}{c}\text{Information}\\ \text{obtained from } \beta\end{array}\right|$$

$$\phi(\mu(A \cap B)) = \phi(\mu(A)\mu(B)) \quad = \quad \phi(\mu(A)) \quad + \quad \phi(\mu(B))$$

This indicates we should take $\phi(t) = -\log t$.

Definition

The information function of α is

$$I(\alpha)(x) = -\sum_{A \in \alpha} \chi_A(x) \log \mu(A).$$

It is natural to assume that if α, β are independent, then:

Information obtained by knowing which element of $\alpha \vee \beta$ we are in	$=$	Information obtained from α	$+$	Information obtained from β
$\phi(\mu(A \cap B)) = \phi(\mu(A)\mu(B))$	$=$	$\phi(\mu(A))$	$+$	$\phi(\mu(B))$

This indicates we should take $\phi(t) = -\log t$.

Definition

The information function of α is

$$I(\alpha)(x) = -\sum_{A \in \alpha} \chi_A(x) \log \mu(A).$$

The entropy of α is the average amount of information:

$$H(\alpha) = \int I(\alpha)\, d\mu = -\sum_{A \in \alpha} \mu(A) \log \mu(A).$$

Conditional information & entropy

Conditional information and entropy are useful generalisations of $I(\alpha)$, $H(\alpha)$.

Conditional information & entropy

Conditional information and entropy are useful generalisations of $I(\alpha)$, $H(\alpha)$.
Let $\mathcal{A}$ be a sub-σ-algebra.

Conditional information & entropy

Conditional information and entropy are useful generalisations of $I(\alpha)$, $H(\alpha)$.

Let $\mathcal{A}$ be a sub-σ-algebra.

For example: if β is a partition then the set of all unions of elements of β is a σ-algebra (also denoted by β).

Conditional information & entropy

Conditional information and entropy are useful generalisations of $I(\alpha)$, $H(\alpha)$.

Let $\mathcal{A}$ be a sub-σ-algebra.

For example: if β is a partition then the set of all unions of elements of β is a σ-algebra (also denoted by β).

How much information do we gain by knowing which element of α we are in, given we know which element of β we are in?

Recall conditional expectation:

$$\mathbb{E}(\cdot \mid \mathcal{A}) : L^1(X, \mathcal{B}, \mu) \longrightarrow L^1(X, \mathcal{A}, \mu).$$

Recall conditional expectation:

$$\mathbb{E}(\cdot \mid \mathcal{A}) : L^1(X, \mathcal{B}, \mu) \longrightarrow L^1(X, \mathcal{A}, \mu).$$

$\mathbb{E}(f \mid \mathcal{A})$ is determined by

1. $\mathbb{E}(f \mid \mathcal{A})$ is $\mathcal{A}$-measurable,
2. $\int_A \mathbb{E}(f \mid \mathcal{A})\, d\mu = \int_A f\, d\mu \ \forall A \in \mathcal{A}$.

Recall conditional expectation:

$$\mathbb{E}(\cdot \mid \mathcal{A}) : L^1(X, \mathcal{B}, \mu) \longrightarrow L^1(X, \mathcal{A}, \mu).$$

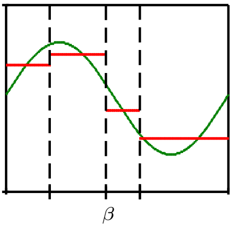
$\mathbb{E}(f \mid \mathcal{A})$ is determined by

1. $\mathbb{E}(f \mid \mathcal{A})$ is $\mathcal{A}$-measurable,
2. $\int_A \mathbb{E}(f \mid \mathcal{A})\, d\mu = \int_A f\, d\mu \ \forall A \in \mathcal{A}$.

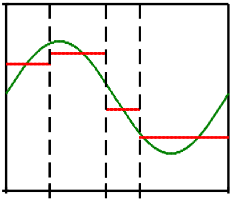
$\mathbb{E}(f|\mathcal{A})$ is the best $\mathcal{A}$-measurable approximation to f.

Consider the σ-algebra β generated by a partition β.

Consider the σ-algebra β generated by a partition β.



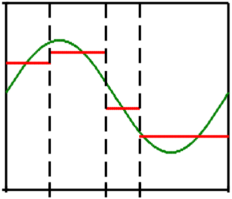
Consider the σ-algebra β generated by a partition β.



β

$$\mathbb{E}(f \mid \beta)(x) = \sum_{B \in \beta} \chi_B(x) \frac{\int_B f \, d\mu}{\mu(B)}.$$

Consider the σ-algebra β generated by a partition β.

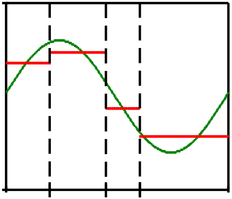


$$\mathbb{E}(f \mid \beta)(x) = \sum_{B \in \beta} \chi_B(x) \frac{\int_B f \, d\mu}{\mu(B)}.$$

Let $A \in \mathcal{B}$. The conditional probability of A given a sub-σ-algebra $\mathcal{A}$ is

$$\mu(A|\mathcal{A}) = \mathbb{E}(\chi_A|\mathcal{A}).$$

Consider the σ-algebra β generated by a partition β.



$$\mathbb{E}(f \mid \beta)(x) = \sum_{B \in \beta} \chi_B(x) \frac{\int_B f \, d\mu}{\mu(B)}.$$

Let $A \in \mathcal{B}$. The conditional probability of A given a sub-σ-algebra $\mathcal{A}$ is

$$\mu(A|\mathcal{A}) = \mathbb{E}(\chi_A|\mathcal{A}).$$

Note:

$$\mu(A \mid \beta) = \sum_{B \in \beta} \chi_B(x) \frac{\mu(A \cap B)}{\mu(B)}.$$

Definition

The conditional information of α given $\mathcal{A}$ is

$$I(\alpha \mid \mathcal{A})(x) = -\sum_{A \in \alpha} \chi_A(x) \log \mu(A \mid \mathcal{A})(x).$$

Definition

The conditional information of α given $\mathcal{A}$ is

$$I(\alpha \mid \mathcal{A})(x) = -\sum_{A \in \alpha} \chi_A(x) \log \mu(A \mid \mathcal{A})(x).$$

The conditional entropy of α given $\mathcal{A}$ is

$$H(\alpha \mid \mathcal{A}) = \int I(\alpha \mid \mathcal{A})\, d\mu.$$

Definition

The conditional information of α given $\mathcal{A}$ is

$$I(\alpha \mid \mathcal{A})(x) = -\sum_{A\in\alpha} \chi_A(x) \log \mu(A \mid \mathcal{A})(x).$$

The conditional entropy of α given $\mathcal{A}$ is

$$H(\alpha \mid \mathcal{A}) = \int I(\alpha \mid \mathcal{A})\, d\mu.$$

The basic identities:

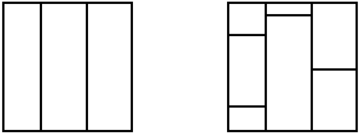
$$\begin{aligned} I(\alpha \vee \beta \mid \gamma) &= I(\alpha \mid \gamma) + I(\beta \mid \alpha \vee \gamma) \\ H(\alpha \vee \beta \mid \gamma) &= H(\alpha \mid \gamma) + H(\beta \mid \alpha \vee \gamma) \end{aligned}$$

Definition

β is a *refinement* of α if every element of α is a union of elements of β.

Definition

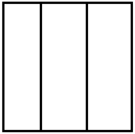
β is a *refinement* of α if every element of α is a union of elements of β.



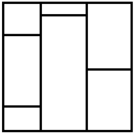
α β

Definition

β is a *refinement* of α if every element of α is a union of elements of β.



α

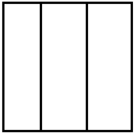


β

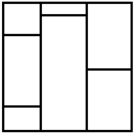
Write $\alpha \leq \beta$.

Definition

β is a *refinement* of α if every element of α is a union of elements of β.



α



β

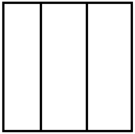
Write $\alpha \leq \beta$.

Facts

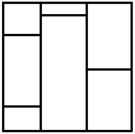
1. $$\beta \leq \gamma \implies \begin{aligned} I(\alpha \vee \beta \mid \gamma) &= I(\alpha \mid \gamma) \\ H(\alpha \vee \beta \mid \gamma) &= H(\alpha \mid \gamma) \end{aligned}$$

Definition

β is a *refinement* of α if every element of α is a union of elements of β.



α



β

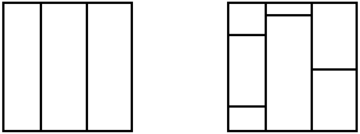
Write $\alpha \leq \beta$.

Facts

$$
\begin{array}{llcl}
1. & \beta \leq \gamma & \Longrightarrow & I(\alpha \vee \beta \mid \gamma) = I(\alpha \mid \gamma) \\
 & & & H(\alpha \vee \beta \mid \gamma) = H(\alpha \mid \gamma) \\
2. & \beta \leq \alpha & \Longrightarrow & I(\beta \mid \gamma) \leq I(\alpha \mid \gamma) \\
 & & & H(\beta \mid \gamma) \leq H(\alpha \mid \gamma)
\end{array}
$$

Definition

β is a *refinement* of α if every element of α is a union of elements of β.



α $\quad$ β

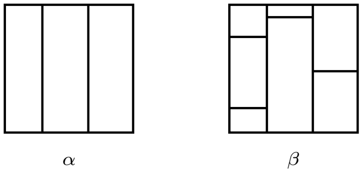
Write $\alpha \leq \beta$.

Facts

$$
\begin{array}{llcc}
1. & \beta \leq \gamma & \Longrightarrow & I(\alpha \vee \beta \mid \gamma) = I(\alpha \mid \gamma) \\
 & & & H(\alpha \vee \beta \mid \gamma) = H(\alpha \mid \gamma) \\
2. & \beta \leq \alpha & \Longrightarrow & I(\beta \mid \gamma) \leq I(\alpha \mid \gamma) \\
 & & & H(\beta \mid \gamma) \leq H(\alpha \mid \gamma) \\
3. & \beta \leq \gamma & \Longrightarrow & H(\alpha \mid \beta) \geq H(\alpha \mid \gamma)
\end{array}
$$

Definition

β is a *refinement* of α if every element of α is a union of elements of β.



Write $\alpha \leq \beta$.

Facts

$$
\begin{array}{llll}
1. & \beta \leq \gamma & \Longrightarrow & I(\alpha \vee \beta \mid \gamma) = I(\alpha \mid \gamma) \\
 & & & H(\alpha \vee \beta \mid \gamma) = H(\alpha \mid \gamma) \\
2. & \beta \leq \alpha & \Longrightarrow & I(\beta \mid \gamma) \leq I(\alpha \mid \gamma) \\
 & & & H(\beta \mid \gamma) \leq H(\alpha \mid \gamma) \\
3. & \beta \leq \gamma & \Longrightarrow & H(\alpha \mid \beta) \geq H(\alpha \mid \gamma)
\end{array}
$$

1,2 follow from the basic identities, 3 follows from Jensen's ineq.

Entropy of an mpt relative to a partition

We can now start to define the entropy $h(T)$ of an mpt T.
We first define the entropy of T relative to a partition.
We need the following:

Entropy of an mpt relative to a partition

We can now start to define the entropy $h(T)$ of an mpt T.
We first define the entropy of T relative to a partition.
We need the following:

Subadditive lemma

Suppose $a_n \in \mathbb{R}$ is subadditive: $a_{n+m} \leq a_n + a_m$.

Then $\lim_{n\to\infty} \frac{a_n}{n}$ exists and equals $\inf_n \frac{a_n}{n}$ (could be $-\infty$).

Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be an mpt.
Let α be a finite or countable partition.

Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be an mpt.
Let α be a finite or countable partition.
Define $T^{-1}\alpha = \{T^{-1}A \mid A \in \alpha\}$ - a countable partition.

Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be an mpt.
Let α be a finite or countable partition.
Define $T^{-1}\alpha = \{T^{-1}A \mid A \in \alpha\}$ - a countable partition.
Note:

$$\begin{aligned}
H(T^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(T^{-1}A) \log \mu(T^{-1}A) \\
&= -\sum_{A\in\alpha} \mu(A) \log \mu(A) \\
&= H(\alpha).
\end{aligned}$$

Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be an mpt.
Let α be a finite or countable partition.
Define $T^{-1}\alpha = \{T^{-1}A \mid A \in \alpha\}$ - a countable partition.
Note:

$$\begin{aligned} H(T^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(T^{-1}A) \log \mu(T^{-1}A) \\ &= -\sum_{A\in\alpha} \mu(A) \log \mu(A) \\ &= H(\alpha). \end{aligned}$$

Define

$$H_n(\alpha) = H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right).$$

Then

$$H_{n+m}(\alpha) \quad = \quad H\left(\bigvee_{j=0}^{n+m-1} T^{-j}\alpha\right)$$

Then

$$
\begin{aligned}
H_{n+m}(\alpha) \quad &= \quad H\left(\bigvee_{j=0}^{n+m-1} T^{-j}\alpha\right) \\
&\overset{\substack{\text{basic}\\ \text{identity}}}{=} \quad H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha \,\middle|\, \bigvee_{j=0}^{n-1} T^{-j}\alpha\right)
\end{aligned}
$$

Then

$$
\begin{aligned}
H_{n+m}(\alpha) &= H\left(\bigvee_{j=0}^{n+m-1} T^{-j}\alpha\right) \\
&\overset{\substack{\text{basic}\\ \text{identity}}}{=} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha \,\middle|\, \bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha\right)
\end{aligned}
$$

Then

$$
\begin{aligned}
H_{n+m}(\alpha) &= H\left(\bigvee_{j=0}^{n+m-1} T^{-j}\alpha\right) \\
&\overset{\substack{\text{basic}\\ \text{identity}}}{=} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha \middle| \bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha\right) \\
&= H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(T^{-n}\bigvee_{j=0}^{m-1} T^{-j}\alpha\right)
\end{aligned}
$$

Then

$$
\begin{aligned}
H_{n+m}(\alpha) &= H\left(\bigvee_{j=0}^{n+m-1} T^{-j}\alpha\right) \\
&\overset{\substack{\text{basic}\\ \text{identity}}}{=} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha \,\middle|\, \bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha\right) \\
&= H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(T^{-n}\bigvee_{j=0}^{m-1} T^{-j}\alpha\right) \\
&= H_n(\alpha) + H_m(\alpha).
\end{aligned}
$$

Then

$$
\begin{aligned}
H_{n+m}(\alpha) \quad &= \quad H\left(\bigvee_{j=0}^{n+m-1} T^{-j}\alpha\right) \\
&\overset{\substack{\text{basic}\\\text{identity}}}{=} \quad H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha \,\middle|\, \bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \\
&\leq \quad H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=n}^{n+m-1} T^{-j}\alpha\right) \\
&= \quad H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(T^{-n}\bigvee_{j=0}^{m-1} T^{-j}\alpha\right) \\
&= \quad H_n(\alpha) + H_m(\alpha).
\end{aligned}
$$

Hence $H_n(\alpha)$ is subadditive.

$\square$

By the subadditive lemma, we can define

$$h_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right)$$

By the subadditive lemma, we can define

$$h_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) = \text{entropy of } T \text{ relative to } \alpha.$$

By the subadditive lemma, we can define

$$h_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) = \text{entropy of } T \text{ relative to } \alpha.$$

Remarks

1. By sub-additivity, $H_n(\alpha) \leq nH(\alpha)$. Hence $0 \leq h_\mu(T,\alpha) \leq H(\alpha)$.

By the subadditive lemma, we can define

$$h_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) = \text{entropy of } T \text{ relative to } \alpha.$$

Remarks

1. By sub-additivity, $H_n(\alpha) \le nH(\alpha)$. Hence $0 \le h_\mu(T,\alpha) \le H(\alpha)$.
2. Using the basic identities and the Increasing Martingale Theorem, we can obtain the following alternative formula for $h_\mu(T,\alpha)$:

$$h_\mu(T,\alpha) = H\left(\alpha \mid \bigvee_{j=1}^{\infty} T^{-j}\alpha\right)$$

By the subadditive lemma, we can define

$$h_\mu(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) = \text{entropy of } T \text{ relative to } \alpha.$$

Remarks

1. By sub-additivity, $H_n(\alpha) \le nH(\alpha)$. Hence $0 \le h_\mu(T,\alpha) \le H(\alpha)$.
2. Using the basic identities and the Increasing Martingale Theorem, we can obtain the following alternative formula for $h_\mu(T,\alpha)$:

$$h_\mu(T,\alpha) = H\left(\alpha \mid \bigvee_{j=1}^{\infty} T^{-j}\alpha\right) = \lim_{n\to\infty} H\left(\alpha | \bigvee_{j=1}^{n} T^{-j}\alpha\right).$$

'Entropy = average amount of information from the present, given the past'

Entropy of an mpt

Let T be an mpt of a probability space $(X, \mathcal{B}, \mu)$.

Entropy of an mpt

Let T be an mpt of a probability space $(X, \mathcal{B}, \mu)$.
Then the entropy of T is:

$$h_\mu(T) = \sup \left\{ h_\mu(T, \alpha) \;\middle|\; \begin{array}{c} \alpha \text{ is a finite or countable} \\ \text{partition s.t. } H(\alpha) < \infty \end{array} \right\}$$

Entropy of an mpt

Let T be an mpt of a probability space $(X, \mathcal{B}, \mu)$.
Then the entropy of T is:

$$h_\mu(T) = \sup \left\{ h_\mu(T, \alpha) \,\middle|\, \begin{array}{c} \alpha \text{ is a finite or countable} \\ \text{partition s.t. } H(\alpha) < \infty \end{array} \right\}$$

Potential problem: working from the definitions, this quantity seems impossible to calculate!

Generators and Sinai's Theorem

Let T be an mpt of the probability space $(X, \mathcal{B}, \mu)$.

Generators and Sinai's Theorem

Let T be an mpt of the probability space $(X, \mathcal{B}, \mu)$.

Definition

A finite or countable partition α is a *generator* for T if T is invertible and

$$\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha \nearrow \mathcal{B}$$

Generators and Sinai's Theorem

Let T be an mpt of the probability space $(X, \mathcal{B}, \mu)$.

Definition

A finite or countable partition α is a *generator* for T if T is invertible and

$$\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha \nearrow \mathcal{B}$$

(i.e. $\mathcal{B}$ is the smallest σ-algebra that contains all elements of all the partitions $\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha$).

Generators and Sinai's Theorem

Let T be an mpt of the probability space $(X, \mathcal{B}, \mu)$.

Definition

A finite or countable partition α is a *generator* for T if T is invertible and

$$\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha \nearrow \mathcal{B}$$

(i.e. $\mathcal{B}$ is the smallest σ-algebra that contains all elements of all the partitions $\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha$).
We say that α is a *strong generator* if

$$\bigvee_{j=0}^{n-1} T^{-j}\alpha \nearrow \mathcal{B}.$$

Remark

To check whether a partition α is a strong generator (resp. generator) it is sufficient to show it separates μ-a.e. pair of points:

Remark

To check whether a partition α is a strong generator (resp. generator) it is sufficient to show it separates μ-a.e. pair of points: for μ-a.e. $x, y \in X$, $\exists n$ s.t. x, y are in different elements of the partition $\bigvee_{j=0}^{n-1} T^{-j}\alpha$ (resp. $\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha$).

Remark

To check whether a partition α is a strong generator (resp. generator) it is sufficient to show it separates μ-a.e. pair of points: for μ-a.e. $x, y \in X$, $\exists n$ s.t. x, y are in different elements of the partition $\bigvee_{j=0}^{n-1} T^{-j}\alpha$ (resp. $\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha$).

Recall:

$$h_\mu(T) = \sup h_\mu(T, \alpha)$$

where the supremum is taken over all partitions of finite entropy.

Remark

To check whether a partition α is a strong generator (resp. generator) it is sufficient to show it separates μ-a.e. pair of points: for μ-a.e. $x, y \in X$, $\exists n$ s.t. x, y are in different elements of the partition $\bigvee_{j=0}^{n-1} T^{-j}\alpha$ (resp. $\bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha$).

Recall:

$$h_\mu(T) = \sup h_\mu(T, \alpha)$$

where the supremum is taken over all partitions of finite entropy.

Sinai's theorem tells us that this supremum is acheived when α is a generator or a strong generator.

Theorem: (Sinai)

Let α be a finite or countable partition with $H(\alpha) < \infty$.

Theorem: (Sinai)

Let α be a finite or countable partition with $H(\alpha) < \infty$. Suppose either:

- T is invertible and α is a generator, or
- α is a strong generator.

Theorem: (Sinai)

Let α be a finite or countable partition with $H(\alpha) < \infty$. Suppose either:

- T is invertible and α is a generator, or
- α is a strong generator.

Then $h_\mu(T) = h_\mu(T, \alpha)$.

Theorem: (Sinai)

Let α be a finite or countable partition with $H(\alpha) < \infty$. Suppose either:

- T is invertible and α is a generator, or
- α is a strong generator.

Then $h_\mu(T) = h_\mu(T, \alpha)$.

This allows us to calculate the entropy of many of our favourite examples.

Example: Markov measures for shifts of finite type

We work in the one-sided setting. Analogous results hold for the two-sided case.

Example: Markov measures for shifts of finite type

We work in the one-sided setting. Analogous results hold for the two-sided case.

Let A be an aperiodic $k \times k$ matrix with corresponding one-sided shift of finite type Σ_A^+. Let σ be the shift map.

Example: Markov measures for shifts of finite type

We work in the one-sided setting. Analogous results hold for the two-sided case.
Let A be an aperiodic $k \times k$ matrix with corresponding one-sided shift of finite type Σ_A^+. Let σ be the shift map.
Let $P = (P_{ij})$ be a stochastic matrix compatible with A.

Example: Markov measures for shifts of finite type

We work in the one-sided setting. Analogous results hold for the two-sided case.
Let A be an aperiodic $k \times k$ matrix with corresponding one-sided shift of finite type Σ_A^+. Let σ be the shift map.
Let $P = (P_{ij})$ be a stochastic matrix compatible with A.
Let $p = (p_1 \ldots, p_k)$ be the unique probability left-eigenvector:

$$pP = p.$$

Example: Markov measures for shifts of finite type

We work in the one-sided setting. Analogous results hold for the two-sided case.
Let A be an aperiodic $k \times k$ matrix with corresponding one-sided shift of finite type Σ_A^+. Let σ be the shift map.
Let $P = (P_{ij})$ be a stochastic matrix compatible with A.
Let $p = (p_1 \ldots, p_k)$ be the unique probability left-eigenvector:

$$pP = p.$$

Recall that the Markov measure μ_P is defined on cylinder sets by:

$$\mu_P[i_o, \ldots, i_{n-1}] = p_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}}.$$

Example: Markov measures for shifts of finite type

We work in the one-sided setting. Analogous results hold for the two-sided case.
Let A be an aperiodic $k \times k$ matrix with corresponding one-sided shift of finite type Σ_A^+. Let σ be the shift map.
Let $P = (P_{ij})$ be a stochastic matrix compatible with A.
Let $p = (p_1 \ldots, p_k)$ be the unique probability left-eigenvector:

$$pP = p.$$

Recall that the Markov measure μ_P is defined on cylinder sets by:

$$\mu_P[i_o, \ldots, i_{n-1}] = p_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}}.$$

Let $\alpha = \{[1], \ldots, [k]\}$ denote the partition of Σ_A^+ into cylinders of length 1.

Easy check: $$H(\alpha) \quad < \quad \infty$$

$$\begin{array}{rrcl} \text{Easy check:} & H(\alpha) & < & \infty \\ \text{Easy check:} & \alpha_n = \bigvee_{j=0}^{n-1} \sigma^{-j}\alpha & = & \{[i_o, \ldots, i_{n-1}]\} \\ & & = & \text{the partition of } \Sigma_A^+ \\ & & & \text{ into cylinders of length } n \end{array}$$

$$\begin{array}{lrcl} \text{Easy check:} & H(\alpha) & < & \infty \\ \text{Easy check:} & \alpha_n = \bigvee_{j=0}^{n-1} \sigma^{-j}\alpha & = & \{[i_o, \ldots, i_{n-1}]\} \\ & & = & \text{the partition of } \Sigma_A^+ \\ & & & \text{ into cylinders of length } n \end{array}$$

Hence α is a strong generator, as α_n separates points.

$$\begin{array}{lrcl}
\text{Easy check:} & H(\alpha) & < & \infty \\
\text{Easy check:} & \alpha_n = \bigvee_{j=0}^{n-1} \sigma^{-j}\alpha & = & \{[i_o, \dots, i_{n-1}]\} \\
 & & = & \text{the partition of } \Sigma_A^+ \\
 & & & \text{ into cylinders of length } n
\end{array}$$

Hence α is a strong generator, as α_n separates points.
Hence we can apply Sinai's theorem:

$$H\left(\bigvee_{j=o}^{n-1} \sigma^{-j}\alpha\right)$$

$$= \quad -\sum_{i_o,\dots,i_{n-1}} \mu[i_0, \dots, i_{n-1}] \log \mu[i_0, \dots, i_{n-1}]$$

$$\begin{aligned} \text{Easy check:} \quad & H(\alpha) &<& \infty \\ \text{Easy check:} \quad & \alpha_n = \bigvee_{j=0}^{n-1} \sigma^{-j}\alpha &=& \{[i_o, \ldots, i_{n-1}]\} \\ && =& \text{ the partition of } \Sigma_A^+ \\ &&& \text{ into cylinders of length } n \end{aligned}$$

Hence α is a strong generator, as α_n separates points.
Hence we can apply Sinai's theorem:

$$\begin{aligned} & H\left(\bigvee_{j=o}^{n-1} \sigma^{-j}\alpha\right) \\ & = \quad - \sum_{i_o,\ldots,i_{n-1}} \mu[i_0, \ldots, i_{n-1}] \log \mu[i_0, \ldots, i_{n-1}] \\ & = \quad - \sum_{i_o,\ldots,i_{n-1}} p_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}} \log(p_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}}) \end{aligned}$$

$$\begin{array}{rrcl} \text{Easy check:} & H(\alpha) & < & \infty \\ \text{Easy check:} & \alpha_n = \bigvee_{j=0}^{n-1} \sigma^{-j}\alpha & = & \{[i_o, \ldots, i_{n-1}]\} \\ & & = & \text{the partition of } \Sigma_A^+ \\ & & & \text{ into cylinders of length } n \end{array}$$

Hence α is a strong generator, as α_n separates points.
Hence we can apply Sinai's theorem:

$$H\left(\bigvee_{j=o}^{n-1} \sigma^{-j}\alpha\right)$$

$$= \quad -\sum_{i_o,\ldots,i_{n-1}} \mu[i_0, \ldots, i_{n-1}] \log \mu[i_0, \ldots, i_{n-1}]$$

$$= \quad -\sum_{i_o,\ldots,i_{n-1}} p_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}} \log(p_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}})$$

$$\overset{\text{re-arranging}}{=} \quad -\sum_i p_i \log p_i - (n-1) \sum_{i,j} p_i P_{ij} \log P_{ij}.$$

Hence

$$h_\mu(\sigma) \overset{\text{Sinai}}{=} h_\mu(\sigma, \alpha)$$

Hence

$$\begin{aligned} h_\mu(\sigma) &\overset{\text{Sinai}}{=} h_\mu(\sigma,\alpha) \\ &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} \sigma^{-j}\alpha\right) \end{aligned}$$

Hence

$$\begin{aligned} h_\mu(\sigma) &\overset{\text{Sinai}}{=} h_\mu(\sigma, \alpha) \\ &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} \sigma^{-j}\alpha\right) \\ &= -\sum_{i,j} p_i P_{ij} \log P_{ij}. \end{aligned}$$

Remark

If μ is the Bernoulli-$(p_1, \ldots, p_k)$ measure then

$$h_\mu(\sigma) = -\sum_i p_i \log p_i.$$

Hence

$$
\begin{aligned}
h_\mu(\sigma) &\overset{\text{Sinai}}{=} h_\mu(\sigma,\alpha) \\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} \sigma^{-j}\alpha\right) \\
&= -\sum_{i,j} p_i P_{ij} \log P_{ij}.
\end{aligned}
$$

Remark

If μ is the Bernoulli-$(p_1, \ldots, p_k)$ measure then

$$h_\mu(\sigma) = -\sum_i p_i \log p_i.$$

If μ is the Bernoulli-$(1/k, \ldots, 1/k)$ measure then

$$h_\mu(\sigma) = \log k.$$

Example

We can model a language (written in the Roman alphabet) as a shift on 26 symbols (corresponding to the 26 letters in the alphabet) with an appropriate Markov measure.

Example

We can model a language (written in the Roman alphabet) as a shift on 26 symbols (corresponding to the 26 letters in the alphabet) with an appropriate Markov measure.

For English:

$P_{\mathbf{QU}}$ should be near 1 as a **Q** is highly likely to be followed by **U**

$P_{\mathbf{FZ}}$ should be near 0 as **F** is unlikely to be followed by **Z**.

Example

We can model a language (written in the Roman alphabet) as a shift on 26 symbols (corresponding to the 26 letters in the alphabet) with an appropriate Markov measure.

For English:

$P_{\mathbf{QU}}$ should be near 1 as a **Q** is highly likely to be followed by **U**

$P_{\mathbf{FZ}}$ should be near 0 as **F** is unlikely to be followed by **Z**.

Experimentally, one can estimate

$$h(\text{English}) = 1.6$$

Example

We can model a language (written in the Roman alphabet) as a shift on 26 symbols (corresponding to the 26 letters in the alphabet) with an appropriate Markov measure.

For English:

$P_{\mathbf{QU}}$	should be near 1 as a **Q** is highly likely to be followed by **U**
$P_{\mathbf{FZ}}$	should be near 0 as **F** is unlikely to be followed by **Z**.

Experimentally, one can estimate

$$h(\text{English}) = 1.6$$

Note that the Bernoulli $(\frac{1}{26}, \ldots, \frac{1}{26})$-measure has entropy $\log 26 = 4.7$.

Example

We can model a language (written in the Roman alphabet) as a shift on 26 symbols (corresponding to the 26 letters in the alphabet) with an appropriate Markov measure.

For English:

$P_{\mathbf{QU}}$ should be near 1 as a **Q** is highly likely to be followed by **U**

$P_{\mathbf{FZ}}$ should be near 0 as **F** is unlikely to be followed by **Z**.

Experimentally, one can estimate

$$h(\mathrm{English}) = 1.6$$

Note that the Bernoulli $(\frac{1}{26}, \ldots, \frac{1}{26})$-measure has entropy $\log 26 = 4.7$.

This suggests that there is a lot of redundancy in English (good for error-correcting!). See Shannon's book on Information Theory.

Entropy as an invariant

Recall that two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *(measure-theoretically) isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

Entropy as an invariant

Recall that two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *(measure-theoretically) isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$.

Entropy as an invariant

Recall that two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *(measure-theoretically) isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \phi \downarrow & & \downarrow \phi \\ Y & \xrightarrow[S]{} & Y \end{array}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$.

Entropy is invariant under isomorphism:

Entropy as an invariant

Recall that two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ are *(measure-theoretically) isomorphic* if there exists a bimeasurable bijection $\phi : X \to Y$ such that

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "S"'] & Y \end{tikzcd}$$

commutes (up to sets of measure zero) and $\mu \circ \phi^{-1} = m$.
Entropy is invariant under isomorphism:

Theorem
If T, S are isomorphic then $h_\mu(T) = h_m(S)$.

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$. Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$H_\mu(\phi^{-1}\alpha) \;=\; -\sum_{A\in\alpha} \mu(\phi^{-1}A)\log\mu(\phi^{-1}A)$$

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$\begin{aligned} H_\mu(\phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(\phi^{-1}A) \log \mu(\phi^{-1}A) \\ &= -\sum_{A\in\alpha} m(A) \log m(A) = H_m(\alpha). \end{aligned}$$

Proof.
Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$\begin{aligned} H_\mu(\phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(\phi^{-1}A)\log\mu(\phi^{-1}A) \\ &= -\sum_{A\in\alpha} m(A)\log m(A) = H_m(\alpha). \end{aligned}$$

More generally

$$H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}(\phi^{-1}\alpha)\right)$$

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$\begin{aligned} H_\mu(\phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(\phi^{-1}A)\log\mu(\phi^{-1}A) \\ &= -\sum_{A\in\alpha} m(A)\log m(A) = H_m(\alpha). \end{aligned}$$

More generally

$$H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}(\phi^{-1}\alpha)\right) = H_\mu\left(\phi^{-1}\bigvee_{j=0}^{n-1} S^{-j}\alpha\right)$$

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$\begin{aligned} H_\mu(\phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(\phi^{-1}A) \log \mu(\phi^{-1}A) \\ &= -\sum_{A\in\alpha} m(A) \log m(A) = H_m(\alpha). \end{aligned}$$

More generally

$$H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}(\phi^{-1}\alpha)\right) = H_\mu\left(\phi^{-1}\bigvee_{j=0}^{n-1} S^{-j}\alpha\right) = H_m\left(\bigvee_{j=0}^{n-1} S^{-j}\alpha\right).$$

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$\begin{aligned} H_\mu(\phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(\phi^{-1}A)\log\mu(\phi^{-1}A) \\ &= -\sum_{A\in\alpha} m(A)\log m(A) = H_m(\alpha). \end{aligned}$$

More generally

$$H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}(\phi^{-1}\alpha)\right) = H_\mu\left(\phi^{-1}\bigvee_{j=0}^{n-1} S^{-j}\alpha\right) = H_m\left(\bigvee_{j=0}^{n-1} S^{-j}\alpha\right).$$

Hence $h_\mu(T, \phi^{-1}\alpha) = h_m(S, \alpha)$.

Proof.

Let α be a finite or countable partition of Y with $H_m(\alpha) < \infty$.
Then $\phi^{-1}\alpha = \{\phi^{-1}A \mid A \in \alpha\}$ is a partition of X.
Note that

$$\begin{aligned} H_\mu(\phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu(\phi^{-1}A)\log\mu(\phi^{-1}A) \\ &= -\sum_{A\in\alpha} m(A)\log m(A) = H_m(\alpha). \end{aligned}$$

More generally

$$H_\mu\left(\bigvee_{j=0}^{n-1} T^{-j}(\phi^{-1}\alpha)\right) = H_\mu\left(\phi^{-1}\bigvee_{j=0}^{n-1} S^{-j}\alpha\right) = H_m\left(\bigvee_{j=0}^{n-1} S^{-j}\alpha\right).$$

Hence $h_\mu(T, \phi^{-1}\alpha) = h_m(S, \alpha)$. Hence $h_\mu(T) = h_m(S)$. $\square$

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ.

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ. Let $\sigma : \Sigma_2 \to \Sigma_2$ be the full one-sided 2-shift with the Bernoulli-$(1/2, 1/2)$ measure μ.

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ. Let $\sigma : \Sigma_2 \to \Sigma_2$ be the full one-sided 2-shift with the Bernoulli-$(1/2, 1/2)$ measure μ.
Define $\phi : \Sigma_2 = \{(x_j)_{j=0}^{\infty} \mid x_j \in \{0, 1\}\} \to [0, 1]$ by

$$\phi(x_0, x_1, \ldots) = \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}}.$$

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ. Let $\sigma : \Sigma_2 \to \Sigma_2$ be the full one-sided 2-shift with the Bernoulli-$(1/2, 1/2)$ measure μ.
Define $\phi : \Sigma_2 = \{(x_j)_{j=0}^{\infty} \mid x_j \in \{0,1\}\} \to [0,1]$ by

$$\phi(x_0, x_1, \ldots) = \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}}.$$

Then

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ. Let $\sigma : \Sigma_2 \to \Sigma_2$ be the full one-sided 2-shift with the Bernoulli-$(1/2, 1/2)$ measure μ.
Define $\phi : \Sigma_2 = \{(x_j)_{j=0}^{\infty} \mid x_j \in \{0,1\}\} \to [0,1]$ by

$$\phi(x_0, x_1, \ldots) = \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}}.$$

Then

- $\phi\sigma = T\phi$,

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ. Let $\sigma : \Sigma_2 \to \Sigma_2$ be the full one-sided 2-shift with the Bernoulli-$(1/2, 1/2)$ measure μ.
Define $\phi : \Sigma_2 = \{(x_j)_{j=0}^{\infty} \mid x_j \in \{0,1\}\} \to [0,1]$ by

$$\phi(x_0, x_1, \ldots) = \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}}.$$

Then

- $\phi\sigma = T\phi$,
- ϕ is a bijection, except on the countable set of points which have non-unique base 2 expansions,

Example: the doubling map and the full 2-shift

Let $Tx = 2x \bmod 1$ be the doubling map with Lebesgue measure λ. Let $\sigma : \Sigma_2 \to \Sigma_2$ be the full one-sided 2-shift with the Bernoulli-$(1/2, 1/2)$ measure μ.
Define $\phi : \Sigma_2 = \{(x_j)_{j=0}^{\infty} \mid x_j \in \{0,1\}\} \to [0,1]$ by

$$\phi(x_0, x_1, \ldots) = \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}}.$$

Then

- $\phi\sigma = T\phi$,
- ϕ is a bijection, except on the countable set of points which have non-unique base 2 expansions,
- $\lambda = \mu\phi^{-1}$ (clear on dyadic intervals, follows for all sets by the Kolmogorov Extension Theorem).

Hence $Tx = 2x \bmod 1$ with Lebesgue measure λ and the full one-sided 2-shift σ with the Bernoulli-$(1/2, 1/2)$ measure μ are isomorphic.

Hence $Tx = 2x \bmod 1$ with Lebesgue measure λ and the full one-sided 2-shift σ with the Bernoulli-$(1/2, 1/2)$ measure μ are isomorphic.

Hence

$$h_\lambda(T) = \log 2 = h_\mu(\sigma).$$

How complete an invariant is entropy?

Given two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ with the same entropy, is it necessarily true that they are isomorphic?

How complete an invariant is entropy?

Given two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ with the same entropy, is it necessarily true that they are isomorphic?

In general, the answer is no.

How complete an invariant is entropy?

Given two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ with the same entropy, is it necessarily true that they are isomorphic?

In general, the answer is no.

However, for two-sided aperiodic shifts of finite type equipped with a Bernoulli or Markov measure, then the answer is yes

How complete an invariant is entropy?

Given two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ with the same entropy, is it necessarily true that they are isomorphic?

In general, the answer is no.

However, for two-sided aperiodic shifts of finite type equipped with a Bernoulli or Markov measure, then the answer is yes

Theorem (Ornstein)

2-sided Bernoulli shifts with the same entropy are isomorphic.

How complete an invariant is entropy?

Given two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ with the same entropy, is it necessarily true that they are isomorphic?

In general, the answer is no.

However, for two-sided aperiodic shifts of finite type equipped with a Bernoulli or Markov measure, then the answer is yes

Theorem (Ornstein)

2-sided Bernoulli shifts with the same entropy are isomorphic.

Theorem (Ornstein and Friedman)

2-sided aperiodic Markov shifts with the same entropy are isomorphic.

How complete an invariant is entropy?

Given two mpts $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$, $S : (Y, \mathcal{A}, m) \to (Y, \mathcal{A}, m)$ with the same entropy, is it necessarily true that they are isomorphic?

In general, the answer is no.

However, for two-sided aperiodic shifts of finite type equipped with a Bernoulli or Markov measure, then the answer is yes

Theorem (Ornstein)

2-sided Bernoulli shifts with the same entropy are isomorphic.

Theorem (Ornstein and Friedman)

2-sided aperiodic Markov shifts with the same entropy are isomorphic.

(The one-sided case is far more subtle.)

Bernoulli systems

Being isomorphic to a Bernoulli shift is a useful and desirable property for a mpt to possess.

Bernoulli systems

Being isomorphic to a Bernoulli shift is a useful and desirable property for a mpt to possess.

Definition

A mpt T of a probability space $(X, \mathcal{B}, \mu)$ is *Bernoulli* if it is isomorphic to a shift σ with some Bernoulli-$(p_1, \ldots, p_k)$ measure.

Bernoulli systems

Being isomorphic to a Bernoulli shift is a useful and desirable property for a mpt to possess.

Definition

A mpt T of a probability space $(X, \mathcal{B}, \mu)$ is *Bernoulli* if it is isomorphic to a shift σ with some Bernoulli-$(p_1, \ldots, p_k)$ measure.

Example

We have already seen that the doubling map with Lebesgue measure is Bernoulli.

Bernoulli systems

Being isomorphic to a Bernoulli shift is a useful and desirable property for a mpt to possess.

Definition

A mpt T of a probability space $(X, \mathcal{B}, \mu)$ is *Bernoulli* if it is isomorphic to a shift σ with some Bernoulli-$(p_1, \ldots, p_k)$ measure.

Example

We have already seen that the doubling map with Lebesgue measure is Bernoulli.

In general, a mpt that exhibits some form of 'hyperbolicity' is, when equipped with a suitable measure, Bernoulli.

Bernoulli systems

Being isomorphic to a Bernoulli shift is a useful and desirable property for a mpt to possess.

Definition

A mpt T of a probability space $(X, \mathcal{B}, \mu)$ is *Bernoulli* if it is isomorphic to a shift σ with some Bernoulli-$(p_1, \ldots, p_k)$ measure.

Example

We have already seen that the doubling map with Lebesgue measure is Bernoulli.

In general, a mpt that exhibits some form of 'hyperbolicity' is, when equipped with a suitable measure, Bernoulli.

For example, hyperbolic toral automorphisms are Bernoulli.

Next lecture

Next lecture

Entropy has been defined in a purely measure-theoretic setting.

There is a topological analogue in the setting of continuous transformations of compact metric spaces: topological entropy.

We will define this and study the connections between measure-theoretic and topological entropy.