INTRODUCTION TO ENTROPY

WOONG CHOI

Abstract. In this paper, I will introduce basic definitions of entropy and talk about the main theorems related to entropy. The main theorems covered in this paper are entropy of the identity transformation, all periodic transformations, and applications of Kolomogrov-Sinai theorem. Basic knowledge in measure theory is assumed.

1. Introduction

Entropy is an important measurement in the field of information theory. Entropy was first developed by Shannon is 1948 from the motivation of limits on compressibility data. Entropy has ties with coding theories, LP hierarchies, and quantum computation. Intuitively, entropy of a random variable X or a set of random variables can be thought of as one of the following:

- The amount of randomness in X
- Minimum number of bits needed to generate a draw from X
- Average number of bits needed to store a draw from X
- Minimum number of bits needed for one to communicate one draw from X with another

For a random variable X, the entropy $H(X)$ is defined as below:

Definition 1.1. $H(X) = \sum_{X \in range(X)} p_X(x) \log_2(1/p_X(x))$.

As an example, if we have four balls in a hat that we can draw with

$$X = \begin{cases} Red & 1/2 \\ Green & 1/4 \\ Blue & 1/8 \\ Orange & 1/8 \end{cases}$$

$H(X) = 1/2 * 1 + 1/4 * (2) + 1/8 * 3 + 1/8 * 3 = 1.75$. We can choose to interpret this entropy value in different ways that are listed in the beginning of this section.

2. Preliminaries

In this section, I will introduce the basic definitions of partition and entropy.

Definition 2.1. A partition of a measure space $(X, \mathcal{D}, \mu)$ is a pairwise disjoint collection of sets $\{A_1, A_2, \ldots\}$ such that $\bigcup_k A_k = X$.

Definition 2.2. The join of the partitions $\mathcal{A} = \{A_1, A_2, \ldots\}$ and $\mathcal{B} = \{B_1, B_2, \ldots\}$ is $\mathcal{A} \vee \mathcal{B} = \{A_i \cap B_j\}$.

Proposition 2.3. *The join of two partitions is another partition.*

Date: March 2024.

Proof. Let $\mathcal{C} = \{(A_i \cap B_j) \cap (A_k \cap B_l)\}$ where either $i \neq k$ or $j \neq l$ holds, but not both. A and B are partitions of sets X and Y, respectively. $(A_i \cap B_j) \cap (A_k \cap B_l) = (A_i \cap A_k) \cap (B_j \cap B_l) = \emptyset \cap \emptyset = \emptyset$. Thus, the elements of the join of two partitions is also pairwise disjoint. ■

The join notation can be extended to multiple sets $\bigvee_{i=1}^{n} \mathcal{A}_i = \{\mathcal{A}_{i_1} \cap \cdots \cap \mathcal{A}_{i_n}\}$. It's also clear that the elements in this join is pairwise disjoint by the same logic as shown in the proof of proposition 1.3. In this paper, the joined partitions will be a set of iterations of T.

Definition 2.4. The entropy of a partition $\mathcal{A} = \{A_1, A_2, \ldots, A_k\}$ is

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

.

Example. Let $(X, \mathcal{D}, \mu)$ be a probability space. Let X be the interval $(0, 1)$. Let partition $\mathcal{A} = \{(0, 1/4), 1/4, (1/4, 1)\}$. $H(\mathcal{A}) = -1/4 \log(1/4) - 0 \log(0) - 3/4 \log(3/4) = 1/4 \log(4) + 3/4 \log(4/3)$.

We can assume that $0 \log(0) = 0$ in the example above because $\lim_{x \to 0} x \log(x) = 0$. In a probability space, entropy is always nonnegative because $\log(\mu(A_i)) \leq 0$ as $0 \leq \mu(A_i) \leq 1$. Notice that if A_i is the full set or null set, then there is no change in entropy, meaning that the term $\mu(A_i) \log \mu(A_i)$ is zero. Also, if $A_i = X$, the all other sets must be empty sets as a partition is a set of pairwise disjoint elements. Alternatively, entropy can be defined for a partition and transformation.

Definition 2.5. The entropy of a finite partition $\mathcal{A}$ and transformation T is

$$H(\mathcal{A}, T) = \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{A})\right)$$

We will not observe the proof of why this limit exists, but the main idea of the proof is that the function $-x \log x$ is bounded by the maximum value of -0.16 approximately. Then, we use that to show that the limit above should exist. Notice that this definition of entropy of a partition uses the previous definition of entropy of a partition on the partition formed by repeatedly applying T. $H(\bigvee_{i=0}^{n-1} T^{-i}(A))$ represents the amount of information after n applications of T on the partition A. $1/n$ in front of H causes the definition of $H(A, T)$ to be the average amount of information added per application of T on A. To remove the dependency on partition, we can define entropy on T as shown below.

Definition 2.6. The entropy of a transformation T is $H(T) = \sup(H(A, T))$, over all finite partitions $\mathcal{A}$.

For a simple application of this definition, we will look into the proof regarding the entropy of the identity transformation and the periodic transformation in the next section.

3. The Identity Transformation and The Periodic Transformation

Proposition 3.1. *Entropy of the identity transformation T is zero.*

Proof. For a finite partition $\mathcal{A}$, $\bigvee_{i=1}^{\infty} \mathcal{A} = \mathcal{A}$. Because entropy of a set of full measure is zero, entropy of an identity transformation T is $lim_{n \to \infty} \frac{H(A)}{n} = 0$ for any partition $\mathcal{A}$, and $\sup(H(A, T)) = 0$. So, we can conclude that the entropy of the identity transformation is 0. ■

For further application of the definition of entropy of a transformation, we will look into periodic functions.

Definition 3.2. A transformation T is periodic if $T(x) = T^m(x)$.

Theorem 3.3. *The Entropy of a periodic transformation is 0.*

Proof. Because T is periodic, $T(x) = T^m(x)$. Then, for a partition $\mathcal{A}$,

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

for all $n > m$. This is because for all $i = k$ with the condition that $k > m$, $T^{-k} = T^{-k+m}$. Because T^{-i} repeats the sets that already exist in the join of the partitions, the entropy does not increase further when $i > m$. For any partition A_i,$\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{m} T^{-i}(\mathcal{A})\right) = 0$, and the same limit holds true for the supremum of all finite partitions. Hence, we can conclude that the entropy of all periodic transformations is zero. ■

4. Applications of Kolmogorov-Sinai Theorem

Definition 4.1. Given a transformation T and a partition $\mathcal{A} = \{A_n\}_{n=1}^{\infty}$, $T(\mathcal{A})$ is equivalent to $\{T(A_n)\}_{n=1}^{\infty}$.

Proposition 4.2. *If T is invertible,then $T(\mathcal{A})$ is also a partition*

Proof. T is invertible, so T is bijective. As T is bijective, $T(A_i) \cap T(A_j) = \emptyset$ for $i \neq j$ because $A_i \cap A_j = \emptyset$, and the transformation applied to each of set also doesn't have any intersection by the definition of a bijective function. ■

Definition 4.3. A T-generator for some invertible T and a measure space $(X, \mathcal{D}, \mu)$ is a partition $\mathcal{A}$ such that $\mathcal{D}$ is generated by $\bigvee_{i=-\infty}^{\infty} T^i(\mathcal{A})$.

Theorem 4.4. *(Kolmogorov-Sinai)If a partition $\mathcal{A}$ is a T-generator, then $H(\mathcal{A}, T) = H(T)$.*

In this paper, we will look at the applications of Kolmogorov-Sinai theorem. The proof can be found in one of the references.

Example. Rotations of the circle have the entropy of zero.

A rational rotation, is a periodic transformation as for any rational number p/q with $(p, q) = 1$, $T^q(x) = T(x)$. Thus, a rational rotation has the entropy of zero. For an irrational rotation, let $\mathcal{A} = \{(0, \pi], [\pi, 2\pi)\}$be the partition of a circle into two arcs. Because each arc has two endpoints, only two new arcs can be formed after an application of T. From reference [1], the maximal entropy of n set partition is $\log(n)$, so $H(T) = H(\mathcal{A}, T) \leq \lim_{n\to\infty} \log(2n + 2)/n = 0$, so the entropy of a rotation of the circle is zero.

References

[1] JACOB FIEDLER. Ergodic theory and entropy.
[2] Ryan O'Donnell. Lecture 20: Information theory. pages 1–2, 2013.
[1] [2]