Information Theory: Entropy

Herbert Ilhan Tanujaya

A0144892W

Definition

Let X be a discrete random variable with probability distribution $p(x)$.

Entropy H_X

$$H_X = -\sum_{x \in X} p(x) \log_2 p(x) = \mathbb{E} \log_2 \left[\frac{1}{p(X)} \right].$$

Here we define $0 \log_2 0 = 0$.
In a sense, the entropy of a random variable shows how "surprising" the event is.

Examples

Entropy of a fair coin

A fair coin has entropy $\frac{1}{2}\log_2(2) + \frac{1}{2}\log_2(2) = 1$.

Entropy of a fair m-sided die

A m-sided die has entropy $\log_2(m)$.

Entropy of n fair coin flips

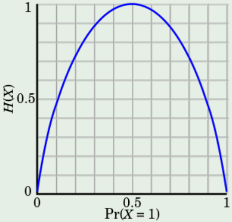
Flipping n coins produce 2^n uniformly distributed possibilities, and hence, the entropy is n.

Examples

Entropy of a Bernoulli trial

If X is a random variable taking values between 0 and 1, where $p(0) = p$ and $p(1) = 1 - p$, its entropy is

$$H_p = -p \log_2 p - (1 - p) \log_2(1 - p).$$



Example

Entropy of an unfair dice

An unfair dice with four faces and

$$p(1) = 1/2, p(2) = 1/4, p(3) = 1/8, p(4) = 1/8$$

has entropy $7/4$, smaller than the one of the corresponding fair dice 2. (This dice is less surprising than the fair dice.)

Motivation [TODO]

Define a function H that takes in a random variable and outputs an integer, such that:

- If a random variable X takes n values, then $H(X)$ is maximized if X is uniform.
- Entropy is additive, in the sense that: if X takes x_i with probability p_i, Y takes x_i with probability q_i, and Z takes x_i with probability $\alpha p_i + \beta q_i$ for all i, then

$$H(Z) = \alpha H(X) + \beta H(Y).$$

Then $H(X) = -\sum p_i \log p_i$ is the only possible function (up to the base of the logarithm).

Properties (TPM 15.7.1)

1. $H(X) \leq \log |S|$.
2. $H(X) \leq H(X, Y) \leq H(X) + H(Y)$.
3. $H(X|Y, Z) \leq H(X|Y)$.

Entropy is subadditive (TPM 15.7.2)

Let $X = (X_1, \ldots, X_n)$ be a random variable taking values in the set $S = S_1 \times S_2 \times \cdots \times S_n$, where each of the coordinates X_i of X is a random variable taking values in S_i. Then

$$H(X) \leq \sum_{i=1}^{n} H(X_i).$$

(This is just induction on the property $H(X, Y) \leq H(X) + H(Y)$.)

TPM 15.7.3

Let $\mathcal{F}$ be a family of subsets of $\{1, 2, \ldots, n\}$ and let p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then

$$|\mathcal{F}| \leq 2^{\sum H(p_i)}.$$

Proof

Let $X = (X_1, \ldots, X_n)$ taking $F \in \mathcal{F}$ with equal probability. Then $H(X) \leq \sum H(X_i)$ implies $\log |\mathcal{F}| \leq \sum H(p_i)$.

TPM 15.7.4

Let $X = (X_1, \ldots, X_n)$ taking values in $S = S_1 \times \cdots \times S_n$, where each X_i takes values in S_i. For an index set $I \subseteq \{1, 2, \ldots, n\}$ let $X(I)$ denote $(X_i)_{i \in I}$. If $\mathcal{G}$ is a family of subsets of $\{1, \ldots, n\}$ and each $i \in \{1, \ldots, n\}$ belongs to at least k members of $\mathcal{G}$, then

$$kH(X) \leq \sum_{G \in \mathcal{G}} H(X(G)).$$

Proof

Use induction. If there is $G \in \mathcal{G}$ where $G = \{1, \ldots, n\}$ we are done. Otherwise, we prove

$$H(X(G \cup G')) + H(X(G \cap G')) \leq H(X(G)) + H(X(G')).$$

TPM 15.7.5

Let $\mathcal{F}$ be a family of vectors in $S_1 \times S_2 \times \cdots \times S_n$. Let $\mathcal{G} = \{G_1, G_2, \ldots, G_m\}$ be a collection of subsets of $N = \{1, 2, \ldots, n\}$, and suppose that each element $i \in N$ belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$ let $\mathcal{F}_i$ be the set of all projections of the members of $\mathcal{F}$ on G_i. Then

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

Proof

Let $X = (X_1, \ldots, X_n)$ taking $F \in \mathcal{F}$ with equal probability. Then $kH(X) \leq \sum H(X(G_i))$ implies $k \log |\mathcal{F}| \leq \sum \log |\mathcal{F}_i|$.