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 "uncertain" the event is.

Examples

Entropy of a fair coin

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

Examples

Entropy of a fair coin

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

Entropy of a fair m-sided die

A m-sided die has entropy $\log m$.

Examples

Entropy of a fair coin

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

Entropy of a fair m-sided die

A m-sided die has entropy $\log m$.

Entropy of n fair coin flips

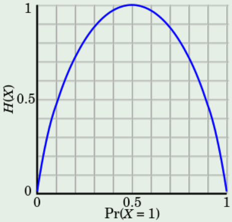
Flipping n coins produce 2^n uniformly distributed possibilities, and hence, the entropy is n.

Examples

Entropy of a Bernoulli trial

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

$$H(p) = -p \log p - (1 - p) \log(1 - p).$$



Examples

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, which is 2. (This dice is less uncertain than the fair dice.)

Motivation: quantifying *information*

Define an information function, I, in terms of an event i with probability p_i, as follows:

Motivation: quantifying *information*

Define an information function, I, in terms of an event i with probability p_i, as follows:

1. $I(p)$ is continuous and decreasing

Motivation: quantifying *information*

Define an information function, I, in terms of an event i with probability p_i, as follows:

1. $I(p)$ is continuous and decreasing
2. $I(p) \geq 0$ and $I(1) = 0$

Motivation: quantifying *information*

Define an information function, I, in terms of an event i with probability p_i, as follows:

1. $I(p)$ is continuous and decreasing
2. $I(p) \geq 0$ and $I(1) = 0$
3. $I(p_1 p_2) = I(p_1) + I(p_2)$, i.e. information due to independent events is additive

Motivation: quantifying *information*

Define an information function, I, in terms of an event i with probability p_i, as follows:

1. $I(p)$ is continuous and decreasing
2. $I(p) \geq 0$ and $I(1) = 0$
3. $I(p_1 p_2) = I(p_1) + I(p_2)$, i.e. information due to independent events is additive

This is solved by $I(p) = -\log p$, up to the base of the logarithm.

Motivation: quantifying *information*

Define an information function, I, in terms of an event i with probability p_i, as follows:

1. $I(p)$ is continuous and decreasing
2. $I(p) \geq 0$ and $I(1) = 0$
3. $I(p_1 p_2) = I(p_1) + I(p_2)$, i.e. information due to independent events is additive

This is solved by $I(p) = -\log p$, up to the base of the logarithm. Now, suppose we have a random variable X, and we sample from it N times. The total amount of information we receive is $\sum N p_i I(p_i)$, and on average, we have $-\sum p_i \log p_i$.

Joint entropy, conditional entropy

Joint entropy

The joint entropy $H(X, Y)$ of a pair of discrete random variables is

$$-\sum \sum p(x, y) \log p(x, y).$$

Joint entropy, conditional entropy

Joint entropy

The joint entropy $H(X, Y)$ of a pair of discrete random variables is

$$-\sum\sum p(x, y) \log p(x, y).$$

Conditional entropy

The conditional entropy $H(Y|X)$ is defined as

$$\sum p(x) H(Y|X = x).$$

Joint entropy, conditional entropy

Joint entropy

The joint entropy $H(X, Y)$ of a pair of discrete random variables is

$$-\sum\sum p(x, y) \log p(x, y).$$

Conditional entropy

The conditional entropy $H(Y|X)$ is defined as

$$\sum p(x) H(Y|X = x).$$

From this definition, we can derive $H(Y|X) = H(X, Y) - H(X)$.

Properties (TPM 15.7.1)

1. $H(X) \leq \log |S|$.

Properties (TPM 15.7.1)

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

Properties (TPM 15.7.1)

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

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

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

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

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

TPM 15.7.3

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

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

Proof

Let $X = (X_1, \ldots, X_n)$ take elements in S with equal probability. Then $H(X) \leq \sum H(X_i)$ implies $\log |S| \leq \sum H(p_i)$.

Kleitman, Shearer, and Sturtevant 1981

Let $\{A_1, A_2, \ldots, A_m\}$ be a collection of k-element sets such that if $\{i, j\} \neq \{p, q\}$, then $A_i \cap A_j \neq A_p \cap A_q$. Then $m = O\left((2-\varepsilon)^k\right)$ for some $\varepsilon > 0$.

Kleitman, Shearer, and Sturtevant 1981

Let $\{A_1, A_2, \ldots, A_m\}$ be a collection of k-element sets such that if $\{i, j\} \neq \{p, q\}$, then $A_i \cap A_j \neq A_p \cap A_q$. Then $m = O\left((2-\varepsilon)^k\right)$ for some $\varepsilon > 0$.

Proof

Let $f(p_i) = \binom{p_i m}{2} / \binom{m}{2}$.

$\log \binom{m}{2} \leq \sum H(f(p_i)) \leq \sum H(p_i^2) \leq k \sum \left(\frac{p_i}{k}\right) \frac{H(p_i^2)}{p_i}$, then use Jensen.

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 $I \subseteq N = \{1, 2, \ldots, n\}$ let $X(I)$ denote $(X_i)_{i \in I}$. If T is a family of subsets of N and each $i \in N$ belongs to at least k members of T, then

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

TPM 15.7.4

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

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

Proof

Use induction. If there is $G \in S$ where $G = N$ 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 (Shearer 1978)

Take $F \subseteq S_1 \times S_2 \times \cdots \times S_n$. Let $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$ be a collection of index sets (i.e. subsets of N), and suppose that each element $i \in N$ belongs to at least k members of $\mathcal{I}$. For each $1 \leq i \leq m$ let F_i be the set of all projections of the members of F on I_i. Then

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

TPM 15.7.5 (Shearer 1978)

Take $F \subseteq S_1 \times S_2 \times \cdots \times S_n$. Let $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$ be a collection of index sets (i.e. subsets of N), and suppose that each element $i \in N$ belongs to at least k members of $\mathcal{I}$. For each $1 \leq i \leq m$ let F_i be the set of all projections of the members of F on I_i. Then

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

Proof

Let $X = (X_1, \ldots, X_n)$ take elements of F with equal probability. Then $kH(X) \leq \sum H(X(I_i))$ implies $k \log |F| \leq \sum \log |F_i|$.

Corollaries of TPM 15.7.5

Take $I_i = N - \{i\}$ for all i, and notice that the volume of a set can be approximated by a collection of fine enough aligned boxes.

TPM 15.7.6

Let B be a measurable body in the n-dimensional Euclidean space, let $Vol(B)$ denote its (n-dimensional) volume, and let $Vol(B_i)$ denote the $(n-1)$-dimensional volume of the projection of B on the hyperplane spanned by all coordinates besides the i-th one. Then

$$(Vol(B))^{n-1} \leq \prod_{i=1}^{n} Vol(B_i).$$

Corollaries of TPM 15.7.5

Take $S_i = \{0, 1\}$ in TPM 15.7.5.

TPM 15.7.7

Let F and $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$ be a collection of subsets of N. Suppose that each element of N belongs to at least k members of $\mathcal{I}$. For each $1 \leq i \leq m$, define $F_i = \{f \cap I_i : f \in F\}$. Then

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

Corollaries of TPM 15.7.5

TPM 15.7.8

Let F be a family of graphs on the labeled set of vertices $\{1, \ldots, t\}$, and suppose that for any two members of F there is a triangle contained in both of them. Then

$$|F| < \frac{1}{4} 2^{\binom{t}{2}}.$$

Corollaries of TPM 15.7.5

TPM 15.7.8

Let F be a family of graphs on the labeled set of vertices $\{1, \ldots, t\}$, and suppose that for any two members of F there is a triangle contained in both of them. Then

$$|F| < \frac{1}{4} 2^{\binom{t}{2}}.$$

Proof

Take N to be the set of all edges of K_t and $\mathcal{I}$ be edges of all possible $K_{\lfloor \frac{t}{2} \rfloor} \cup K_{\lceil \frac{t}{2} \rceil}$.