Information Entropy

Entropy

- A measure of information, uncertainty, randomness, ...
- We have used probability tools to describe uncertainty.
- What is the relation between entropy and probabilities?

Entropy

- Let X be a random variable that takes on finitely many possible values $x_1, \ldots, x_n$ and let $p_1, \ldots, p_n$ be the associated probabilities. The entropy $H(x)$ of X is a number that depends only on the probabilities $p_1, \ldots, p_n$.

$$H(X) = -\underbrace{\sum_{i=1}^{n} p_i \log_2 p_i}$$

Examples

- Tossing a fair coin: Let X be the random variable of the outcomes.

$$P_0 = \frac{1}{2} \qquad P_1 = \frac{1}{2}$$

$$H = -\left(P_0 \log_2 P_0 + P_1 \log_2 P_1\right)$$

$$= -\left(\frac{1}{2} \cdot \log_2 \frac{1}{2} + \frac{1}{2} \cdot \log_2 \frac{1}{2}\right) = -\log_2 \frac{1}{2} = 1$$

Examples

- Tossing a *biased* coin: Let X be the random variable of the outcomes.

$P_0 = p.$ $\quad P_1 = 1-p$

$$H = -p\lg p - (1-p)\lg(1-p)$$

- The entropy function H is continuous in the variables p_i.

Properties of Information Entropy

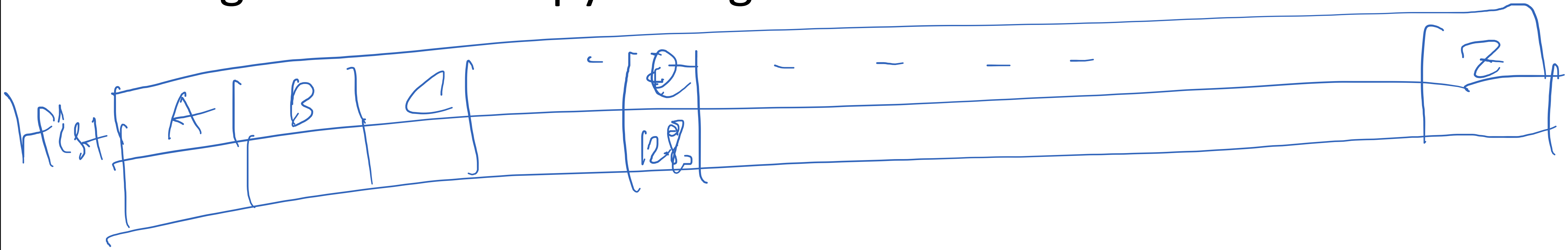
- The entropy of an n-outcome uniformly distributed random variable is always lower than or equal to that of an $(n+1)$-outcome uniformly distributed random variable.

Properties of Information Entropy

- If the *outcomes of* a random variable X are grouped into non-overlapping *events*. The entropy of X is the weighted sum of the entropy for the component groups.

Entropy of English Letters

- Single-letter entropy of English



$$H(L) = -(0.0815 \log_2 0.0815 + \cdots + 0.0008 \log_2 0.0008) \approx 4.132$$

$P_{'a'}$ $\qquad$ $P_{'z'}$

Entropy of English Letters

- Two-letter entropy of English
  - $p(\text{“TH”}) = 0.00315, \quad p(\text{“AN”}) = 0.00172, \quad p(\text{“IT”}) = \cdots$

26^2

$$H(L^2) = -(\cdots + 0.00315 \log_2 0.00315 + \cdots + 0.00172 \log_2 0.00172)\frac{1}{2} \approx \frac{7.12}{2} = 3.56 < 4.12$$

Entropy of English Letters

- Per letter entropy of English

$$H(L) = \lim_{n\to\infty} \frac{H(L^n)}{n} \rightarrow 1.5$$