

# Information Theory

石政修

NCKU

March 25, 2026

## 1. Entropy

A **probability space**  $(\Omega, \mathcal{F}, \mathcal{P})$  is a mathematical model of random experiments.

- **Sample space**  $\Omega$  is the set of all possible outcomes.
- **Event space**  $\mathcal{F}$  is the set of all subsets of  $\Omega$ . It satisfies:
  - $\Omega \in \mathcal{F}$
  - If  $A \in \mathcal{F}$ , then  $A^c \in \mathcal{F}$
  - If  $A_1, A_2, \dots \in \mathcal{F}$ , then  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
- **Probability measure**  $\mathcal{P}$  assigning a probability to each event. It satisfies:
  - $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$
  - $\mathcal{P}(\Omega) = 1$
  - If  $A_1, A_2, \dots \in \mathcal{F}$  are disjoint events, then  $\mathcal{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathcal{P}(A_i)$

# Random Variable

A **random variable**  $X$  on  $(\Omega, \mathcal{F}, \mathcal{P})$  is a real-valued function:

$$X: \Omega \rightarrow \mathbb{R}$$

the range of  $X$  is denoted as  $\mathcal{X}$  ( $\mathcal{X} \subseteq \mathbb{R}$ ).

**Probability mass function** (PMF) is a distribution of  $X$ :

$$P_X(x) = \mathcal{P}(X = x) \quad \text{where } x \in \mathcal{X}$$

*Many properties of a random variable are determined by its distribution, for example, the expected value  $\mathbb{E}[X]$  is determined by  $P_X(x)$ .*

# Surprise Function

In a random experiment, we derived a **surprise function**  $s(A)$  to measure the "surprise" of observing an event  $A \subseteq \Omega$ , and three axioms are naturally show up:

- If  $\mathcal{P}(A) = 1$ , it is unsurprising, that is,  $s(A) = 0$ .
- The less  $\mathcal{P}(A)$  is, the more surprising it is.
- If  $\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$ , then  $s(A \cap B) = s(A) + s(B)$ . That is, the information of independent events is additive.

There is a unique function that meets these three axioms:

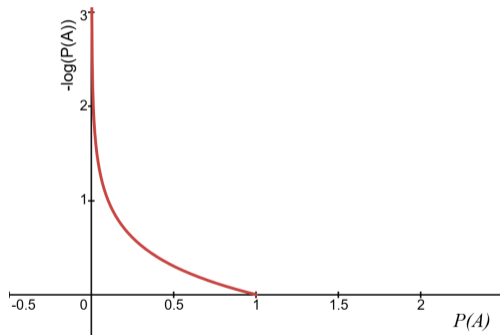
$$s(A) = -k \log(\mathcal{P}(A)) \quad \text{where } k > 0$$

equivalently,

$$s(A) = -\log_b(\mathcal{P}(A)) \quad \text{where } b > 1$$

# Information

Surprise is also called **information content** (or self-information), because less probable events provide more information. For example, a guy win a lottery (low probability) tell us more things than lose a lottery (high probability).



*If  $P(A) = 0$ , then  $s(A) = \infty$ . It means the surprising of an impossible event occurs is infinity.*

# Entropy

The **entropy**  $H_b(X)$  in base  $b$  of  $X$  is the average surprise of the event  $A = \{X = x\}$ :

$$H_b(X) = \mathbb{E}[s(\{X = x\})] = \mathbb{E}[-\log_b(P_X(x))] = - \sum_{x \in \Omega} P_X(x) \log_b(P_X(x))$$

in short,

$$H_b(X) = - \sum_{x \in \mathcal{X}} P_X(x) \log_b(P_X(x))$$

- When  $P_X(x) = 0$ , the convention defines  $0 \times \log 0 = 0$
- For  $b = 2$ ,  $H_2(X) = H(X)$  and  $\log_2 = \log$
- The unit of entropy in base 2 is *bit* or *shannon*, for base  $e$  is *nat* and for base 10 is *hartly*

**The End**