Information theory and coding

Bartek Blaszczyszyn
Notes by Antoine Groudiev

3rd February 2024

Introduction

This document is Antoine Groudiev's class notes while following the class *Théorie de l'information et codage* (Information theory and coding) at the Computer Science Department of ENS Ulm. It is freely inspired by Bartek Blaszczyszyn's class notes.

1 Entropy and source coding

We shall introduce *Shannon's entropy* of a probability distribution on a discrete space and study its basic properties. Our goal is to prove *Shannon's source coding theorem* formulated in 1948. It will allow us to interpret the entropy as a notion of the *amount of information* "carried" by random variables of a given distribution.

1.1 Shannon's entropy

Let $\mathcal{X}$ be a finite or countable set, and $p := \{p(x) \mid x \in \mathcal{X}\}$ be a probability distribution on $\mathcal{X}$.

Definition (Shannon's entropy)**.** We define (Shannon's) entropy $H(p)$ of p to be:

$$H(p) := -\sum_{x \in \mathcal{X}} p(x) \log p(x) \tag{1}$$

with the convention that $0 \log 0 = 0$, and $a \log 0 = -\infty$ for $a > 0$. We will later on discuss the base of the logarithm.

Definition (Entropy of a random variable)**.** Let X be a random variable on $\mathcal{X}$ with distribution p, that is $\mathbb{P}(X = x) = p(x)$, also denoted $X \sim p$. We define:

$$H(X) := H(p) = -\mathbb{E}(\log p(X)) \tag{2}$$

Observe that $0 \leqslant H(p) \leqslant +\infty$, and that $H(p) = 0$ if and only if X is constant almost everywhere.

Property. *Entropy is invariant with respect to deterministic injective mapping $f : \mathcal{X} \to \mathcal{Y}$:*

$$H(X) = H(f(X))$$

The entropy $H(p)$ can be interpreted as the *amount of information* carried on average by one realization from the distribution p. Later in this chapter, we shall prove a result supporting this interpretation.

Definition (Entropy units)**.** The unit of the entropy depends on the *base of the logarithm*:

- In binary basis, when $\log = \log_2$, we denote $H(p) = H_2(p)$, and its unit is the $[bit/symbol]$ (per realization of X).
- In arbitrary basis $b > 0$, when $\log = \log_b$, we denote $H(p) = H_b(p)$, and its unit is the $[b - digit/symbol]$ (a b-digit is a digit which can take b values).
- In basis e, when $\log = \ln$, we denote $H(p) = H_e(p)$, and its unit is the $[nat/symbol]$ (nat is the natural unit of information).

The conversion between units can be done by changing the base of the logarithm:

$$H_b(p) = \frac{H_2(p)}{\log_2(b)}$$

Example (Bernoulli distribution)**.** *Let $\mathcal{X} = \{0, 1\}$, and p the Bernoulli distribution such as*

$$\begin{cases} p(0) = p \\ p(1) = 1 - p \end{cases}$$

Therefore, we have $H(p) = -p\log(p) - (1-p)\log(1-p)$. The Bernoulli distribution with the maximum entropy is:

$$\max_{0 \leqslant p \leqslant 1} H_2(p) = H_2(1/2) = 1\ [bit/symbol]$$

Example (Uniform distribution)**.** *Let $\mathcal{X}$ be a finite set, and p the uniform distribution, that is:*

$$\forall x \in \mathcal{X},\ p(x) := \frac{1}{|\mathcal{X}|}$$

Therefore, we have $H(p) = \log(|X|)$.

Example (Geometric distribution)**.** *Let $\mathcal{X} = \mathbb{N}^\star$ and p the geometric distribution of parameter $p > 0$, that is:*

$$\forall n \in \mathbb{N}^\star,\ p(n) = p(1-p)^{n-1}$$

Recall that $\mathbb{E}[X] = \frac{1}{p}$ when X follows a geometric law of parameter p.

Therefore, we have:

$$H(p) = \log\left(\frac{1-p}{p}\right) - \frac{1}{p}\log(1-p)$$