

Information theory and coding

Bartek Blaszczyzyn
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 Blaszczyzyn'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:

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

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:

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

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

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

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

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:

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

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 $H(p) = \log(|\mathcal{X}|)$.

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

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

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

Therefore, we have:

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

1.2 Gibbs' inequality

Theorem (Gibbs' inequality). Let p and q be two probability distributions on \mathcal{X} . Then:

$$H(p) = - \sum_{x \in \mathcal{X}} p(x) \log p(x) \leq - \sum_{x \in \mathcal{X}} p(x) \log q(x) \quad (3)$$

Moreover, if $H(p) < \infty$, then there is equality in (3) if and only if $p = q$.

The right-hand-side of (3) is called *cross entropy* between p and q .

Proof. Let $x \sim p$. Gibbs' inequality is equivalent to:

$$\mathbb{E}[\log p(X)] \geq \mathbb{E}[\log q(X)]$$

If $\mathbb{E}[\log q(X)] = -\infty$, the inequality is trivial. Otherwise, since we have $\mathbb{E}[\log q(X)] \leq 0$:

$$\begin{aligned}\mathbb{E}[\log q(X)] - \mathbb{E}[\log p(X)] &= \mathbb{E}[\log q(X) - \log p(X)] \\ &= \mathbb{E}\left[\log\left(\frac{q(X)}{p(X)}\right)\right]\end{aligned}$$

\log being concave, by applying Jensen's inequality, we obtain:

$$\begin{aligned}\mathbb{E}\left[\log\left(\frac{q(X)}{p(X)}\right)\right] &\leq \log \mathbb{E}\left[\frac{q(X)}{p(X)}\right] \\ &= \log \sum_{x \in \mathcal{X}} \frac{q(x)}{p(x)} p(x) \\ &= \log \sum_{x \in \mathcal{X}} q(x) \\ &= \log 1 = 0\end{aligned}$$

The equality in Jensen's inequality holds if and only if $\frac{q(X)}{p(X)}$ is almost surely constant, that is $p = \lambda q$ almost surely; furthermore, we must have $\lambda = 1$ since both p and q are distributions, hence $p = q$ almost surely. \square

Corollary (Uniform distribution maximizes entropy). *Let p be a probability distribution on some set \mathcal{X} with $|\mathcal{X}| < \infty$. Then:*

$$0 \leq H(p) \leq \log(|\mathcal{X}|)$$

and the equality holds if and only if p is uniform on \mathcal{X} .

Proof. Let $X \sim p$ and be q the uniform distribution on \mathcal{X} . By Gibbs' inequality:

$$H(p) \leq - \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{1}{|\mathcal{X}|}\right) = \log |\mathcal{X}|$$

Notice that $\log |\mathcal{X}|$ is the entropy of the uniform distribution q . \square

Corollary (Geometric distribution maximizes entropy in the set of probability measures on \mathbb{N}^* having given expectation). *Let p be a probability distribution on $\mathcal{X} = \mathbb{N}^*$ with mean $\mu = \sum_{n \geq 1} np(n) < \infty$. Then:*

$$H(p) \leq (\mu + 1) \log(\mu + 1) - \mu \log \mu$$

where the right-hand-side is the entropy of the geometric distribution with parameter $1/\mu$.

Proof. Let p be a probability distribution on $\mathcal{X} = \mathbb{N}^*$ with mean $\mu < \infty$, and q the geometric distribution of parameter $1/\mu$. According to Gibbs' inequality,

$$\begin{aligned}H(p) &\leq - \sum_{n \geq 1} p(n) \log q(n) \\ &= - \sum_{n \geq 1} p(n) \log \left(\frac{1}{\mu} \left(1 - \frac{1}{\mu}\right)^{n-1} \right) \\ &= \sum_{n \geq 1} p(n) \log \mu - \sum_{n \geq 1} (n-1) p(n) \log \left(1 - \frac{1}{\mu}\right) \\ &= \log \mu - \log \left(1 - \frac{1}{\mu}\right) (\mu - 1) \\ &= \log \mu - (\log(\mu - 1) - \log \mu) (\mu - 1) \\ &= \log \mu + \mu \log \mu - \mu \log(\mu - 1) + \log(\mu - 1) - \log \mu \\ &= \mu \log \mu - (\mu - 1) \log(\mu - 1) = H(q)\end{aligned}$$

\square