

# Information Theory<sup>\*</sup>

Manuel Gijón Agudo

September 2017 - January 2018

---

<sup>\*</sup>Chapter 2 in Cryptography class in Master's degree at UPC

## Índice

<b>1. Basic knowledge of coding</b>	<b>2</b>
1.1. Introduction . . . . .	2
1.2. Uniquely decodeable codes . . . . .	2
1.3. Optimal codes . . . . .	3
1.4. Extension of sources . . . . .	3
<b>2. Introduction</b>	<b>4</b>
<b>3. Properties of the entropy function</b>	<b>4</b>
<b>4. Shannon-Fano Code</b>	<b>4</b>
<b>5. Product of sources</b>	<b>5</b>
<b>6. Markov Chains</b>	<b>5</b>
<b>7. Sources with memory</b>	<b>6</b>

# 1. Basic knowledge of coding

## 1.1. Introduction

**Definition 1:** a **source** is a finite set  $\mathcal{S}$  together with a set of random variables  $(X_1, X_2, \dots)$  whose range is  $\mathcal{S}$ .

If  $P(X_n = \mathcal{S}_i)$  only depends on  $i$  and not on  $n$  then we say the source is **stationary** and if the  $X_n$  are independent then it's **memoryless**.

Insert example here

**Definition 2:** Let  $\mathcal{T}$  be a finite set called **alphabet**. A map  $\mathfrak{C} : \mathbb{S} \longrightarrow \mathbb{U}_{n \geq 1} T^n$  is called a **code**.

If  $|\mathcal{T}| = r$  then  $\mathfrak{C}$  is a  **$r$ -ary code**.

A code extends from  $\mathbb{S}$  to  $T \cup T^2 \cup \dots$  to  $\mathbb{S} \cup \mathbb{S}^2 \cup \dots$  to  $T \cup T^2 \cup \dots$  in obvious way.

insert example here

**Definition 3:** The **average word-length** of a code  $\mathfrak{C}$  is  $L(\mathfrak{C}) := \sum_{i=1}^n p_i l_i$  where  $l_i$  is the length of the image of the symbol of  $\mathbb{S}$ , which is emitted with probability  $p_i$ .

For now, we write  $\mathfrak{C}$  to be the image of  $\mathfrak{C}$ .

## 1.2. Uniquely decodeable codes

**Definition 4:** If for any sequences  $u_1 \dots u_n = v_1 \dots v_m$  in  $\mathfrak{C}$  implies  $m = n$  and  $u_i = v_i$  for  $i = 1, \dots, n$  then we say that  $\mathfrak{C}$  is **uniquely decodeable**.

insert example here

insert example here

insert example here

Let  $\mathfrak{C}_0 = \mathfrak{C}$ :

- $\mathfrak{C}_n := \{\omega \in T \cup T^2 \cup \dots \mid u\omega = v \text{ for some } u \in \mathfrak{C}_{n-1}, v \in \mathfrak{C} \text{ or } u\omega = v \text{ for some } u \in \mathfrak{C}, v \in \mathfrak{C}_{n-1}\}$
- $\mathfrak{C}_\infty := \bigcup_{k \geq 1} \mathfrak{C}_k$

Since everything is finite either  $\mathfrak{C}_m = \emptyset$  for some  $m$  and then  $\mathfrak{C}_n = \emptyset$  for  $n \geq m$  or it will be periodic and start repeating.

**Theorem 1:**  $\mathfrak{C}$  is uniquely decodeable  $\iff \mathfrak{C} \cap \mathfrak{C}_\infty = \emptyset$ .

*proof:* Insert proof here

insert example here

insert example here

insert example here

**Definition 5:** A code is a **prefix-code** if no codeword is prefix of another (ie.  $\mathfrak{C}_1 = \emptyset$ ).

A prefix code is uniquely decodeable.

**Theorem 2: (Kraft's inequality)**  $\exists$   $r$ -ary prefix code with word lengths  $l_1, l_2, \dots, l_q \iff$

$$\sum_{i=1}^q r^{-l_i} \leq 1$$

*proof:* Insert proof here

insert example here

**Theorem 3: (McMillan's inequality)**  $\exists$   $r$ -ary uniquely decodeable code with word lengths  $l_1, l_2, \dots, l_q \iff$

$$\sum_{i=1}^q r^{-l_i} \leq 1$$

*proof:* Insert proof here

### 1.3. Optimal codes

Let be  $\mathbb{S}$  a source with symbols  $s_1, \dots, s_q$  emitted with probabilities  $p_1, \dots, p_q$  and  $\mathfrak{C}$  is a code which encodes  $s_i$  with a codeword length  $l_i$ . Recall  $L(\mathfrak{C}) = \sum_{i=1}^q p_i l_i$ .

**Definition 6:** An **optimal code** for  $\mathbb{S}$  is an uniquely decodeable code  $\mathfrak{D}$  such that  $L(\mathfrak{C}) \geq L(\mathfrak{D})$  for all unique decodeable code  $\mathfrak{C}$ .

inset example here

insert example here

**Definition 7:** A code constructed in this way is called a **Huffman code**.

insert example here

### 1.4. Extension of sources

## 2. Introduction

**Definition 1:** the **information** conveyed by a source is a function  $I : S \rightarrow [0, \infty)$  where  $S$  is a **source**<sup>1</sup> with the properties:

- $I(s_i)$  is a decreasing function of the propability  $p_i$ , with  $I(s_i) = 0$  if  $p_i = 1$ .
- $I(s_i s_j) = I(s_i) + I(s_j)$ , ie.the information gained by two symbols is the sum of the information obtained from each where the source has symbols  $s_1, \dots, s_q$  emitted with probabilities  $p_1, \dots, p_q$ .

**Lemma 1:**  $I(s_i) = -\log_r p_i$  for some  $r$ .

*proof:* Insert proof here

---

**Definition 2:** The  $r$ -ary **entropy**  $H_r(S)$  of a source  $S$  is the average information conveyed by  $S$ .

$$H_r(S) := -\sum_{i=1}^q p_i \log_r p_i$$

, by convection  $x \log_r x$  evaluated at 0 is 0.

Insert five examples

## 3. Properties of the entropy funcion

**Theorem 1:**  $H_r(S) \leq \log_r q$  with equality if and only iff  $S$  is the source where each symbol is emitted with probability  $1/q$ .

*proof:* Insert proof here

---

**Theorem 2:**  $H_r(S) \leq L(C)$  for unique decodeable code  $C$ .

*proof:* Insert proof here

---

## 4. Shannon-Fano Code

Let  $S$  be the source with symbols  $s_i$  and probabilities  $p_i$ . Let  $l_i := \lceil \log_r 1/p_i \rceil$ .

Then:  $\sum_{i=1}^q r^{-l_i} \leq \sum r^{-\log_r 1/p_i} = \sum p_i = 1$

**Definition 3:** by Kraft exists a prefix code with woed length  $l_1, l_2, \dots, l_1$ . This code is called **Shannon-Fano code**.

---

<sup>1</sup>A **source** is a finite set  $S$  together with a sequence of random variables  $X_i$  whose range is  $S$

Inert example here

**Lemma 2:** For the Shannon-Fano code  $C$ :  $H_r(S) \leq L(C) < H_r(S) + 1$ .

*proof:* Insert proof here

---

## 5. Product of sources

Let  $S$  and  $T$  be two memoryless sources,  $S$  with symbols  $s_i$  and probabilities  $p_i$  and  $T$  with symbols  $t_j$  and probabilities  $q_j$ .

**Definition 4:** The **product source**  $S \times T$  is a source with symbols  $s_i t_j$  and probabilities  $p_i q_j$ .

**Theorem 3:**  $H_r(S \times T) = H_r(S) + H_r(T)$ .

*proof:* Insert proof here

---

**Corollary 1:**  $H_r(S^n) = nH_r(S)$ .

**Theorem 4: Noiseless Coding** The average word length  $L_n$  of an optimal code of  $S^n$  satisfies:

$$\frac{L_n}{n} \rightarrow H_r(S), n \rightarrow \infty$$

*proof:* Insert proof here

---

some examples

## 6. Markov Chains

**Definition 4:** A **Markov Chain** is a sequence of random variables where  $X_{n+1}$  depends only for  $X_n$ .

$$P(X_{n+1} = s_j | X_n = s_i) = p_{i,j}$$

This can be represented in a direct graph and also by a matrix  $P := (p)_{i,j}$ .

Suppose  $u_0$  is the vector which describes the initial distribution, ie. the  $i$ -th coordinate of  $u_0$  is probability we start at  $s_i$ . Probability of being in the  $i$ -th state after  $r$  steps is the  $i$ -th coordinate of  $u_0 P^r$ .

**Theorem 5:** if  $\exists r \in \mathbb{N}$  such that  $P^r$  has no zero entries, then  $u_0 P^r \rightarrow u$ , as  $n \rightarrow \infty$ .

**Definition 5:** This vector  $u$  is called the **stationary distribution**. It is normalised eigenvector of  $P^t$  with eigenvalue 1, ie.  $u_j = \sum_i p_{i,j} u_i$  and  $\sum_j u_j = 1$ .

**Definition 6:** If  $P$  is the matrix of a Markov Chain and  $\exists r$  such that  $P^r$  has non zero entries then we say that the Markov Chain is **regular**.

## 7. Sources with memory

Suppose  $S$  is a Markov Chain source with random variables  $X_1, X_2, \dots$  such that

$$P(X_{n+1} = s_j | X_n = s_j) = p_{i,j}$$

**Definition 7:**  $S$  is **not memoryless**, but it is stationary.

**Theorem 6:** suppose  $S$  is a regular Markov Chain source with stationary distribution  $u = (u_1, \dots, u_n)$ . Let  $S'$  be the stationary memoryless source with the same source elements as  $S$  (where  $s_i$  is emitted with probability  $w_i$ ). Then:

$$H_r(S) \leq H_r(S')$$

*proof:* Insert proof here

---