Information Theory & Error Correcting Codes

Written by HADIOUCHE Azouaou.

Disclaimer

This document contains the lectures given by Dr. Seffah.

> To separate the contents of the course to actual additions or out of context information, a black band will be added by its side like the one on this comment.

Contents

Chapter 0

Remainders

1. Congruences & $\mathbb{Z}/m\mathbb{Z}$ Arithmetic

Definition 1.1 (Congruence Of Integers / Congruence Class): *Let $a, b, m \in \mathbb{Z}$ with $m > 0$, we say that a is congruent to b modulo m and we write $a \equiv b \bmod m$ if $m \mid a - b$ which gives an equivalence relation. The class of a in the congruence relation by m is called the congruence class of a modulo m, which is $\overline{a} = a + m\mathbb{Z}$.*

Theorem 1.2: *Let $a, b, c, d \in \mathbb{Z}$ and $m \in \mathbb{N}$, with $a \equiv b \bmod m$ and $c \equiv d \bmod m$ then we have the following statements are true:*

1. $a + c \equiv b + d \bmod m$.
2. $a - c \equiv b - d \bmod m$.
3. $ac \equiv bd \bmod m$.

We denote $\mathbb{Z}/m\mathbb{Z}$, the set of all congruence classes modulo m.

2. Euler φ Function

Definition 2.3 (φ Function): *The Euler φ function is defined as $\varphi : \mathbb{Z} \to \mathbb{N}$, where $\varphi(n) = \#\{x \in [\![0, m-1]\!] \mid \gcd(x, m) = 1\}$.*

Lemma: *Let $r, m, p \in \mathbb{N}$ where p is prime, then $\gcd(m, p^r) \neq 1 \Leftrightarrow p \mid m$*

Proof. $\Rightarrow$ let $d = \gcd(m, p^r)$, since $d \mid p^r$ and p is prime then $d = p^k$ for some $k \in \mathbb{N}$, and $p^k = d \mid m$ thus $p \mid p^k$ and $p^k \mid m$ so $p \mid m$. $\Leftarrow$ suppose that $p \mid m$ and since $p \mid p^r$ too then $p \mid \gcd(m, p^r)$ given $p > 1$ then $\gcd(m, p^r) \neq 1$. □

Proposition 2.4: *Let p be a prime number, then $\varphi(p) = p - 1$ and $\varphi(p^r) = p^r - p^{r-1} = p^r\left(1 - \frac{1}{p}\right)$.*

Proof. Let $p, r \in \mathbb{N}$ with p prime. From the lemma, we have that for $m \in \mathbb{N}$, $\gcd(p^r, m) = 1 \Leftrightarrow p \nmid m$, the numbers that are divisible by p in $[\![0, p^r - 1]\!]$ are of the form kp where $k \in [\![0, p^{r-1} - 1]\!]$, so the number of numbers that are divisible by p in $[\![0, p^r - 1]\!]$ is $p^r - p^{r-1}$ so $\varphi(p^r) = p^r - p^{r-1}$. □

Theorem 2.5 (Lagrange): *Let G be a finite group and H a subgroup of G, then $\# H \mid \# G$.*

This theorem is already done multiple times in algebra so no need to prove it here.

Lemma 2.6: *Let $n \in \mathbb{N}$, $x \in (\mathbb{Z}/n\mathbb{Z})^*$ (invertible) $\Leftrightarrow \gcd(x, n) = 1$.*

Proof. $\Rightarrow$ Suppose that $x \in (\mathbb{Z}/n\mathbb{Z})^*$, then $\exists y \in \mathbb{Z}/n\mathbb{Z}$ such that $xy = 1 \bmod n$, then $xy = 1 + kn \Rightarrow xy - kn = 1$ by Bezout theorem we have $\gcd(x, n) = 1$. $\Leftarrow$ SUppose that $\gcd(x, n) = 1$, by Bezout theorem, $\exists y, k \in \mathbb{Z}, xy + kn = 1$ and thus $xy + kn = 1 \bmod n \Rightarrow xy = 1 \bmod n$ so $x \in (\mathbb{Z}/n\mathbb{Z})^*$. □

We need this lemma since $\varphi(n) = \# (\mathbb{Z}/n\mathbb{Z})^*$.

Theorem 2.7 (Euler): *Let m be a positive integer modulo a be an integer relatively prime to m then $a^{\varphi(m)} \equiv 1 \bmod m$.*

Proof. Let a relatively prime to m, then $\gcd(a, m) = 1 \Rightarrow a \in (\mathbb{Z}/m\mathbb{Z})^*$, a has finite order thus $\exists k \in \mathbb{N}, a^k = 1 \bmod m$, $\langle a \rangle$ is a subgroup of $(\mathbb{Z}/m\mathbb{Z})^*$ so $k \mid \#(\mathbb{Z}/m\mathbb{Z})^* = \varphi(m)$ and thus $\varphi(m) = kb$, thus $a^{\varphi(m)} = a^{kb} = \left(a^k\right)^b = 1^b = 1 \bmod m$, which completes the proof. □

Corollary 2.8 (Fermat): *Let p be a prime, if the integer a is not divisible by p then $a^p \equiv a \bmod p$.*

Proof. Just take $m = p$ in Euler's theorem, then $\varphi(m) = p - 1$ so $a^{\varphi(m)} = a^{p-1} = 1 \bmod m$, so $a^p = a \bmod m$. □

2.1. Quadratic Residues

Definition 2.1.9 (Quadratic Residue): *Let p be an odd prime and a an integer not divisible by p, we say that a is a quadratic residue modulo p if there exists $x \in \mathbb{Z}$, such that $x^2 \equiv a \bmod p$.*

Theorem 2.1.10: *An integer a is a quadratic residue modulo p if and only if $\gcd(a, p) = 1$ and a has a square rest modulo p.*

Proof. $\Rightarrow$ Suppose that a is a quadratic residue modulo p, that is, there exists $x \in \mathbb{Z}$ such that $x^2 \equiv a \bmod p$. If $\gcd(a, p) \neq 1$ then $p \mid a$ so $a \equiv 0 \bmod p$ thus $x = 0$, so necessarily $\gcd(a, p) = 1$ and a has x^2 remainder when divided by p thus the rest is a perfect square. $\Leftarrow$ Suppose now that $\gcd(a, p) = 1$ and a has a square rest modulo p, then $a \not\equiv 0 \bmod p$ and $a \equiv x^2 \bmod p$ thus a is quadratic residue. □

Definition 2.1.11 (Legendre Symbol): *Let p be an odd prime, define the Legendre symbol as:*

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } \gcd(a, p) = 1 \text{ and } a \text{ is a quadratic residue} \bmod p \\ -1 & \text{if } \gcd(a, p) = 1 \text{ and } a \text{ is not a quadratic residue} \\ 0 & \text{if } p \text{ divides } a \end{cases}$$

Theorem 2.1.12: *Let p be an odd prime, for every integer a*

$$\left(\frac{a}{p}\right) = a^{\frac{p-1}{2}} \bmod p$$

Proof. We treat by cases

1. $p \mid a$: then $a = 0 \bmod p \Rightarrow a^{\frac{p-1}{2}} = 0^{\frac{p-1}{2}} = 0 \bmod p$.
2. $p \nmid a$: then $\gcd(a, p) = 1$, by Fermat's little theorem we have that $a^{p-1} \equiv 1 \bmod p$, the square roots of 1 are -1 and 1 only thus $a^{\frac{p-1}{2}} \equiv 1 \text{ or } -1 \bmod p$.
    1. If a is a quadratic residue then $\exists x \in \mathbb{Z}, x^2 \equiv a \bmod p \Rightarrow a^{\frac{p-1}{2}} \equiv x^{2\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \bmod p$.
    2. If $a^{\frac{p-1}{2}} \equiv 1 \bmod p$, we have that $\mathbb{Z}_p$ is a finite field thus there exists a generator g of $\left(\mathbb{Z}_p\right)^*$ (primitive element theorem in finite fields), then $a \equiv g^k \bmod p$, we have then $a^{\frac{p-1}{2}} \equiv g^{k\frac{p-1}{2}} \equiv 1 \bmod p$. Given that the generator g has order $p - 1$ then $k\frac{p-1}{2} \equiv 0 \bmod p - 1$ thus $k\frac{p-1}{2} = m(p - 1) \Rightarrow k = 2m$ so k is even, thus $a \equiv g^{2m} = (g^m)^2 \bmod p$ so a is a quadratic residue modulo p.

Given that a is a quadratic residue modulo p if and only if $a^{\frac{p-1}{2}} \equiv 1 \bmod p$ and $a^{\frac{p-1}{2}} \equiv 1 \text{ or } -1 \bmod p$ then necessarily if a is not a quadratic residue, $a \equiv -1 \bmod p$. □

Exercise 2.1.13:

1. *Decompose into partial fractions in $\mathbb{R}[x]$ the rational function*

$$\frac{x}{x^4 + x^2 + 1}$$

2. *Let K be a commutative field, and let $p : x^2 + \lambda x + \mu$ be a monic polynomial of degree 2, show that p is reducible over K if and only if it has a root in K.*
3. *Let $K = \mathbb{Z}/5\mathbb{Z}$ be the field of residue classes, factor the polynomial $(x^2 + 4)(x^2 + 3)$ into irreducible factors over K.*
4. *Still with $K = \mathbb{Z}/5\mathbb{Z}$, decompose into partial fractions the rational function*

$$\frac{x - \overline{2}}{(x^2 + \overline{4})(x^2 + \overline{3})}$$

1. Decomposing into partial fractions: we have that the denominator decomposes into the following two irreducible polynomials

$$\begin{aligned} x^4 + x^2 + 1 &= x^4 + 2x^2 + 1 - x^2 \\ &= \left(x^2+1\right)^2 - x^2 \\ &= \left(x^2+x+1\right)\left(x^2-x+1\right) \end{aligned}$$

   and thus we have that the fraction decomposes as follows

$$\frac{x}{x^4+x^2+1} = \frac{ax+b}{x^2+x+1} + \frac{cx+d}{x^2-x+1}$$

   by multiplying both sides by $x^4 + x^2 + 1$ we get

$$\begin{aligned} &x = (ax+b)\left(x^2-x+1\right) + (cx+d)\left(x^2+x+1\right) \\ &0 = ax^3 - ax^2 + ax + bx^2 - bx + b + cx^3 + cx^2 + cx + dx^2 + dx + d - x \\ &0 = (a+c)x^3 + (-a+b+c+d)x^2 + (a-b+c+d-1)x + (b+d) \end{aligned}$$

$$\Rightarrow \begin{cases} a+c &= 0 \\ -a+b+c+d &= 0 \\ a-b+c+d-1 &= 0 \\ b+d &= 0 \end{cases} \Rightarrow \begin{cases} a = 0 \\ b = -1/2 \\ c = 0 \\ d = 1/2 \end{cases}$$

   thus we get

$$\frac{x}{x^4+x^2+1} = \frac{1}{2\left(x^2-x+1\right)} - \frac{1}{2\left(x^2+x+1\right)}$$

2. Let K be a commutative field and let $P(x) = x^2 + \lambda x + \mu$. $\Rightarrow$ Suppose that $P(x)$ is reducible, then $P(x) = Q(x) \cdot R(x)$ with $\deg(Q) \neq 0$ and $\deg(R) \neq 0$ and since $\deg(Q) + \deg(R) = \deg(P)$ then necessarily $\deg(Q) = \deg(R) = 1$, so $Q(x) = ax + b$ and $R(x) = cx + d$ and since $a, b \in K$ then $-\frac{a}{b} \in K$ which is a root of Q and thus a root of P so P has a root in K. $\Leftarrow$ Suppose that $P(x)$ has a root α, then $(x - \alpha)$ divides P and thus we get that P is reducible.
3. Let $K = \mathbb{Z}/5\mathbb{Z}$, factor $\left(x^2+4\right)\left(x^2+3\right)$ into irreducible factors over K, $-3 = 2 \bmod 5$ is not a quadratic residue since $2^{\frac{5-1}{2}} = 4 = -1 \bmod 5$ thus $x^2 + 3$ is irreducible given it has no roots. While $-4 = 1$ is a quadratic residue with roots -1 and 1 so $\left(x^2+4\right)\left(x^2+3\right) = \left(x^2-1\right)\left(x^2+3\right) = (x-1)(x+1)\left(x^2+3\right)$.
4. Decomposing now the partial fraction in $\mathbb{Z}/5\mathbb{Z}$,

$$\begin{aligned} \frac{x-2}{\left(x^2+4\right)\left(x^2+3\right)} &= \frac{x-2}{(x+4)(x+1)\left(x^2+3\right)} \\ &= \frac{a}{x+4} + \frac{b}{x+1} + \frac{cx+d}{x^2+3} \end{aligned}$$

   by multiplying all parts by $\left(x^2+4\right)\left(x^2+3\right)$ we get

$$x - 2 = a(x+1)\left(x^2+3\right) + b(x+4)\left(x^2+3\right) + (cx+d)(x+4)(x+1)$$

   we set the following values for x

   - $x = -1$: $-3 = b(-1+4)(1+3) = 12b = 2b \Rightarrow b = 3 * (-3) = -9 = 1$.
   - $x = +1$: $-1 = a(1+1)(1+3) = 8a = 3a \Rightarrow a = 2 * (-1) = -2 = 3$.

Exercise 2.1.14:

1. *Decompose* 561 *into prime factors.*
2. *Let* a *be an indeterminate, for any integer* $n \geq 1$*, justify the formula*

$$a^n - 1 = (a-1)\left(a^{n-1} + a^{n-2} + \cdots + a + 1\right)$$

3. *Justify for any integer* $m \geq 1$ *and any integer* $n \geq 1$ *the formula*

$$a^{mn} - 1 = (a^m - 1)\left(a^{(n-1)m} + a^{(n-2)m} + \cdots + a^m + 1\right)$$

4. *Let* $a \in \mathbb{Z}$ *be an integer, prove* $a^{561} - a = a\left(a^{2 \cdot 280} - 1\right) = a\left(a^2 - 1\right)m$ *for some integer* m*.*
5. *For any* $a \in \mathbb{Z}$*, show that* $a(a-1)(a+1)$ *is divisible by* 3*.*
6. *Show that* $a^{561} - a$ *is a multiple of* 3*.*
7. *Show that* $a^{561} - a$ *is a multiple of* 17*.*
8. *Show that* $a^{561} - a$ *is a multiple of* 11*.*
9. *Show that* $a^{561} \equiv a \bmod 561$*.*

1. The prime decomposition of 561 is $3 \cdot 11 \cdot 17$.
2. Let a be a formal variable, we have

$$(a-1)\sum_{i=0}^{n-1} a^i = \sum_{i=0}^{n-1} a^{i+1} - a^i = \sum_{i=1}^{n} a^i - \sum_{i=0}^{n-1} a^i = a^n - 1$$

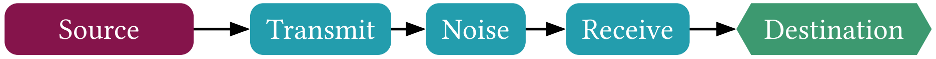
3. By taking $a = a^m$ in the previous inequality we get the result.

4. Clearly $561 = 2 \cdot 280 + 1$ thus $a^{561} - a = a\big(a^{560} - 1\big) = a\big(a^{2 \cdot 280} - 1\big)$ and we have from the previous question by taking $m = 2$ and $n = 280$ that $a^{561} - a = a\big(a^2 - 1\big)\Big(\sum_{i=1}^{279} a^{2i}\Big) = a\big(a^2 - 1\big)m$.
5. Let $a \in \mathbb{Z}$, we have that $a = 3q + r$ with $0 \leq r < 2$ from the Euclidean division algorithm, then $a(a-1)(a+1) = (3q+r)(3q+(r-1))(3(q+1)+(r-2))$ and since $0 \leq r < 2$ then necessarily either r, $r-1$ or $r-2$ is 0, and thus it has one of the factors divisible by 3.
6. $a^{561} - a$ is divisible by 3 since $a^{561} - a = a(a+1)(a-1)m$ which is divisible by 3 from the previous question.
7. $a^{16} = 1 \bmod 17$ by Fermat little theorem, then $a^{560} = a^{16 \cdot 35} = 1 \bmod 17$ thus, $a^{561} = a \bmod 17 \Rightarrow 17 \mid a^{561} - a$.
8. Same here, $a^{10} = 1 \bmod 11$ by Fermat little theorem, then $a^{560} = a^{10 \cdot 56} = 1 \bmod 11$ thus, $a^{561} = a \bmod 11 \Rightarrow 11 \mid a^{561} - a$.
9. Given that all the divisors of 561 divide $a^{561} - a$ then it is divisible by their product so $561 \mid a^{561} - a \Rightarrow a^{561} \equiv a \bmod 561$.

Chapter 1

Introduction To Information Theory

The digital revolution has transformed how we communicate, store and process information. This chapter introduces the fundamental concepts of information theory, a discipline at the intersection of mathematics, computer science and telecommunication.



This diagram present the general structure of any communication system. The communication process begins with an information source.

Notation: *Let Σ be a set:*

- $\Sigma^n = \Sigma \times \cdots \times \Sigma$ *is the set of words of length* n*. We take that* $\Sigma^0 = \{\varepsilon\}$ *where* ε *is the empty word.*
- $\Sigma^* = \bigcup_{i \geq 0} \Sigma^n$ *the set of all possible words over* Σ*.*
- $\Sigma^+ = \bigcup_{i \geq 1} \Sigma^n$ *the set of all possible non-empty words over* Σ*.*
- *For* $x \in \Sigma^*$*, we define* $|x| = \min\{n \in \mathbb{N} \mid x \in \Sigma^n\}$*.*
- *The set of binary letters as* $\mathbb{B} = \{0, 1\}$

Definition 1.1 (Symbol/Message): *Let* $\mathcal{X} = \{x_1, \ldots, x_m\}$ *be a finite set of alphabets. An element of* $\mathcal{X}$ *is called a symbol, and elements of* $\mathcal{X}^+$ *are called a message with alphabet* $\mathcal{X}$*.*

To model information that was given from a message, we consider the fact that a message contains more information the more unpredictable it is. For example, consider a sender that keeps sending the message “Hello!” over and over again, at each iteration, it becomes predictable what the next message would be, thus there is no more information that is passed. Uncertainty is a property of a random process, thus a good way to model a message is to consider it as a random variable X that has values in $\mathcal{X}$. From the previous example too, we can think that the information in a message is related to how uncertain it is, that is, how less probable it would be sent.

Notation: *For the sake of simplifying notation, we will denote* $P(X = x)$ *as* $P(x)$*, same for other random variables, like* $P(Y = y)$ *as* $P(y)$*. Same for all the remaining notations like* $P(y|x)$ *which is* $P(Y = y|X = x)$*.*

Definition 1.2 (Measure Of Information): *Let* X *be a random variable that represents the probability that a message* x *is sent as* $P(X = x) = P(x)$*, we define the measure of information of the message* x *as*

$$I(x) = -\log_2 P(x)$$

$\log_2$ is taken for the usual reason that information is represented in binary, and that if it is represented in any other base, it would have just a linear factor added. In practice, transmission lines tend to cause irregularities in the signal, that is, it alters the contents of the message with some noise. Thus, in transmission lines we are interested in the distribution of the message y that are received, given some sent message x, which will be measured by $P(y|x)$.

1.1. Entropy

After we defined a measure of information, which we discussed to be a measure of uncertainty, we need a way to quantify the average amount of uncertainty for all the values of X. Thus, we define the entropy as follows.

Definition 1.1.3 (Entropy): *Let* X *be a random variable with values in* $\mathcal{X}$*, we define the entropy* $H(X)$ *as the average amount of information, that is*

$$H(x) = \mathbb{E}[I(X)] = -\sum_{x \in X} P(x) \cdot \log_2 P(x)$$

Notice that the information of some message x depends only on its unpredictability, that is, its probability of occurrence $P(x)$ not its value.

Example: Let $\mathcal{X} = \{0, 1\}$, and X a discrete random variable with values in $\mathcal{X}$, which has the distribution

$$P(x) = \begin{cases} 0.9 \text{ if } x = 0 \\ 0.1 \text{ if } x = 1 \end{cases}$$

We obtain that

$$I(x) = \begin{cases} 0.152 \text{ if } x = 0 \\ 3.321 \text{ if } x = 1 \end{cases}$$

. And the measure of average amount of information is $H(X) = 0.4689$.

1.1.1. Source Coding

The most important part of coding theory is achieving the most efficient reliable and secure coding. The first part is the focus of this section, by assigning a code from each symbol in $\mathcal{X}$, we try to achieve the minimum bound possible of letters to send to transmit our message.

> **Definition 1.1.1.4 (Coding Function):** *Let $c : \mathcal{X} \to \mathbb{B}^+$, we call it a coding function, which takes characters of our set of symbols, and represent it a binary string in $\mathbb{B}^+$.*

To measure the efficiency of our coding function, we define the average code length, the less the average, the more efficient the transmission will be.

> **Definition 1.1.1.5 (Average Code Length):** *Let $c : \mathcal{X} \to \mathbb{B}^+$, a coding function, and consider the function*
>
> $$\tau : \{c : \mathcal{X} \to \mathbb{B}^x\} \to \mathbb{R} \qquad c \mapsto \tau_c = \sum_{x \in \mathcal{X}} p(x) \cdot |c(x)|$$

Example: Take a horse race with 8 horses, we want to send a message in binary that indicates which horse has won. And suppose that the probabilities of winning for each horse is as follows

Horse	1	2	3	4	5	6	7	8
Probability Of Winning	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$

If we just send the index of the winner horse, we get the following coding function

Horse i	1	2	3	4	5	6	7	8
$c_1(i)$	000	001	010	011	100	101	110	111

If we do it just blindly, we will need 3 bits to describe all the possible winner horses from $000, 001, 010, \cdots, 110, 111$, calculating the average of description length we get

$$\tau_{c_1} = \sum_{x \in X} 3 \cdot p(x) = 3 \text{ bits}$$

Giving us an average of 3 bits for the transmission to give exactly who horse is the winner. An aspect that we did not use in the previous part is how likely do horses win in this case, which we can use to improve the amount of bits that would be sent on the channel. Consider the distribution of the horses winning as follows

Notice that if we give a smaller message for the horses that are more probable to win, and less to horses that are less likely to win, then we can reduce some of the data that will be used to specify the winner. We take the following encoding

Horse i	1	2	3	4	5	6	7	8
$c_2(i)$	0	10	110	1110	111100	111101	1111110	111111

Now, we recalculate the average description length to get

$$\tau_{c_2} = \sum_{x \in X} |c(x)| \cdot p(x) = 2 \text{ bits}$$

we have reduced the average description length by 1 bit. If we calculate the entropy in this case we get $H(X) = -\sum_{x \in X} P(x) \cdot \log_2 P(x) = 1.83475$, we see that the reduced average code length is more than the entropy.

Theorem 1.1.1.6 (Shannon's First Theorem): *Let X be a random variable with values in $\mathcal{X}$, for any coding function $c : \mathcal{X} \to \mathbb{B}^+$ we have*

$$\tau_c \geq H(X)$$

Proof.

Lemma 1.1.1.7 (Kraft Inequality): *Let X be a random variable with values in $\mathcal{X}$, $c : \mathcal{X} \to \mathbb{B}^+$ a coding function and $l(x) = |c(x)|$*

□