CODING THEORY

Reference

Elements of information theory T.M. Cover and J.A. Thomas Wiley

- Fourier analysis — t-f aspects
- Shannon theory — informational / statistical aspects of random signals

1948

Q1 : what is information content of message ?

Q2 : How much info can we send reliably through a noisy channel ?

1

$\boxed{X} \xrightarrow{p(x)}$ 0110100110 ... $H(x)$ = entropy

source

→ DATA COMPRESSION

↳ Source coding theorem (Shannon 1st theorem)

2

$\xrightarrow{X} \boxed{p(Y|X)} \xrightarrow{Y}$ $C = \max I(X;Y)$ = capacity

channel

→ ERROR-CORRECTING CODE

↳ Channel coding theorem (Shannon 2nd theorem)

CH 1 : ENTROPIES

X = random variable over alphabet $\mathcal{X} = \{x\}$ → symbol

characterized by $p(x) = \mathbb{P}[X = x] \quad \forall x \in \mathcal{X}$

DEF $\boxed{H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)}$ (bit) $p(x) = ?$

↳ l'entropie

- Unbiased coin, $\mathcal{X} = \{\text{heads}, \text{tails}\} = \{0, 1\}$

$$H(X) = -\frac{1}{2}\underbrace{\log_2 \frac{1}{2}}_{\underbrace{-\log_2 2}_{1}} - \frac{1}{2}\log_2 \frac{1}{2} = 1 \text{ bit}$$

- Uniform distribution $\mathcal{X} = \{1, 2, \dots, n\}$ $\quad f(x) = \frac{1}{n} \quad \forall x \in \mathcal{X}$

$$H(X) = -\sum_{x=1}^{n} \frac{1}{n} \underbrace{\log_2 \frac{1}{n}}_{-\log_2 n} = \log_2 n$$

$$H(X) = \log |\mathcal{X}| \longrightarrow \text{cardinality of alphabet } \mathcal{X}$$

- H is functional of $f(x)$ (and not of the value) we can take random values like
  $\{0, 1\}$
  $\{8, 9\}$
  $\{A, B\}$

- $\mathcal{X} = \{\text{heads}, \text{tails}, \text{side}\}$ with probabilities $\frac{1-\varepsilon}{2}$, $\frac{1-\varepsilon}{2}$, ε

$$H = \underbrace{-\frac{1-\varepsilon}{2}\log\frac{1-\varepsilon}{2}}_{-\frac{1}{2}} \underbrace{-\frac{1-\varepsilon}{2}\log\frac{1-\varepsilon}{2}}_{-\frac{1}{2}} - \varepsilon \log \varepsilon$$

$\varepsilon \to 0$

$$\lim_{\varepsilon \to 0} \left(-\varepsilon \log \varepsilon\right) = \lim_{\varepsilon \to 0} \frac{\log \varepsilon}{-\frac{1}{\varepsilon}} = \lim_{\varepsilon \to 0} \frac{1/\varepsilon}{1/\varepsilon^2} = \lim_{\varepsilon \to 0} \varepsilon = 0$$

- $\underline{H(X) \geq 0}$ $\quad 0 \leq f(x) \leq 1 \quad \forall x$

$$H = \mathop{E}_{f(x)}\left(-\log f(x)\right) \geq 0 \quad \forall x$$

Shannon (1948)

<u>3 Axioms</u>: $p(x) = \{p_1, p_2, \dots, p_n\}$

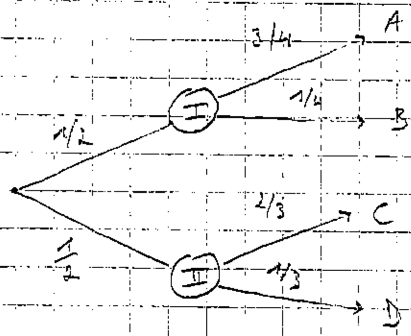
(a) $H(p_1, p_2, \dots, p_n)$ continuous in p_i's $(i = 1, \dots, n)$

(b) if all p_i's are equal $\left(p_i = \frac{1}{n}\right)$ then H should be monotonously increasing with n.

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i = -\sum^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n$$

(c) GROUPING AXIOM

<u>ex:</u>



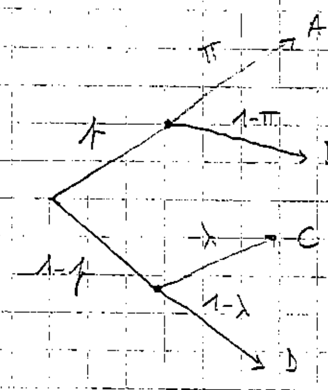
A prob $\frac{3}{8}$

B prob $\frac{1}{8}$

C prob $\frac{1}{3}$

D prob $\frac{1}{6}$

$$H\left[\frac{3}{8}, \frac{1}{8}, \frac{1}{3}, \frac{1}{6}\right] = \underbrace{H\left[\frac{1}{2}, \frac{1}{2}\right]}_{1^{st}\ \text{step}} + \underbrace{\frac{1}{2} H\left[\frac{3}{4}, \frac{1}{4}\right] + \frac{1}{2} H\left[\frac{2}{3}, \frac{1}{3}\right]}_{2^{nd}\ \text{step}}$$



$0 \le p \le 1$

$0 \le \pi \le 1$

$0 \le \lambda \le 1$

$$\begin{aligned}
&H\left[p\pi, p(1-\pi), (1-p)\lambda, (1-p)(1-\lambda)\right] \\
&= -p\pi \log(p\pi) - p(1-\pi)\log\left(p(1-\pi)\right) - (1-p)\lambda \log\left((1-p)\lambda\right) \\
&\quad - (1-p)(1-\lambda)\log\left((1-p)(1-\lambda)\right) \\
&= -\underline{p\pi \log p} - p\pi \log \pi - \underline{p(1-\pi)\log p} - p(1-\pi)\log(1-\pi) \\
&\quad - \underline{(1-p)\lambda \log(1-p)} - (1-p)\lambda \log \lambda - \underline{(1-p)(1-\lambda)\log(1-p)} \\
&\quad - (1-p)(1-\lambda)\log(1-\lambda) \\
&= \underline{-p \log p - (1-p)\log(1-p)} + p\left\{-\pi \log \pi - (1-\pi)\log(1-\pi)\right\} \\
&\quad + (1-\pi)\left\{-\lambda \log \lambda - (1-\lambda)\log(1-\lambda)\right\} \\
&= \underbrace{H\left[p, 1-p\right]}_{} + \underbrace{p \cdot H\left[\pi, 1-\pi\right] + (1-p)\, H\left[\lambda, 1-\lambda\right]}_{}
\end{aligned}$$

Interpretation of $H(X)$

1) Measure of uncertainty
(amount of randomness)

2) Measure of information content

$\boxed{X} \longrightarrow x \longrightarrow$ Nbr of YES/NO questions that need to be answered in order to determine x

ex 1 : $\mathcal{X} = \{A, B, C, D\}$

$p(A) = p(B) = p(C) = p(D) = \frac{1}{4}$ $\qquad H(X) = \log_2 4 = 2$ bits/letter

$\boxed{X} \longrightarrow x \xRightarrow{2 \text{ bits}} \begin{cases} A \to 00 \\ B \to 01 \\ C \to 10 \\ D \to 11 \end{cases}$ $\qquad L = 2$ bits / letter

ex 2 : $\mathcal{X} = \{A, B, C, D\}$

$L = \frac{1}{2}1 + \frac{1}{4}2 + \frac{1}{8}3 + \frac{1}{8}3 = \frac{4+4+3+3}{8} = \frac{7}{4}$ bits/letter

$\frac{7}{4} < 2$

HUFFMAN CODE

x	$p(x)$	Codeword	length
A	1/2	1	1
B	1/4	01	2
C	1/8	001	3
D	1/8	000	3

$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}$

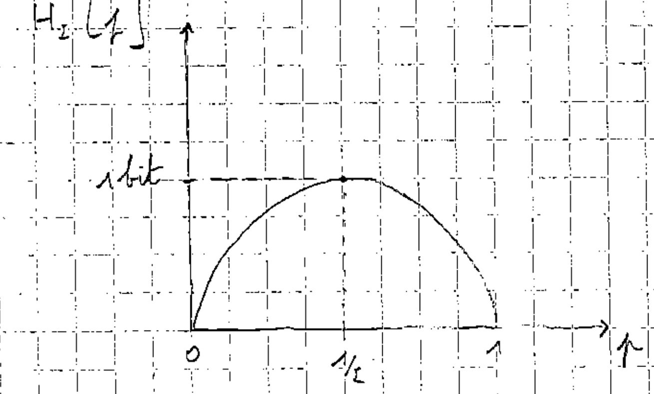
(without considering the codewords)

$\boxed{H(X) \leq L}$ = length of shortest description of random variable X

Entropy of binary variable

$X = \begin{cases} 0 & \text{prob } 1-p \\ 1 & \text{prob } p \end{cases}$ BERNOULLI VARIABLE of parameter p $(0 \leq p \leq 1)$

$$H(X) = -p \log_2 p - (1-p)\log_2(1-p) \equiv H_2[p]$$

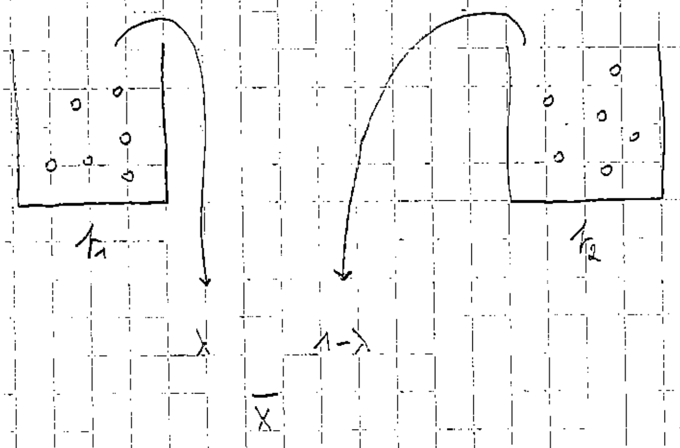


$H_2[p]$ is **concave** funct. of p

$$H_2[\lambda p_1 + (1-\lambda) p_2] \geq \lambda H[p_1] + (1-\lambda) H[p_2]$$

with $0 \leq \lambda \leq 1$ $\qquad H(\bar{X}) \qquad \frac{1}{2} H(X_1) + \frac{1}{2} H(X_2)$

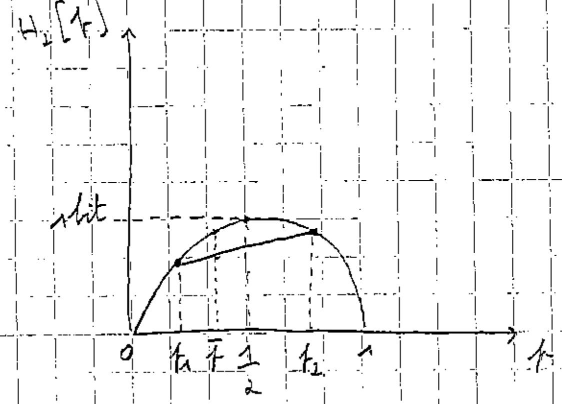
$$X_1 = \begin{cases} 0 & \text{prob } 1-p_1 \\ 1 & \text{prob } p_1 \end{cases} \qquad X_2 = \begin{cases} 0 & \text{prob } 1-p_2 \\ 1 & \text{prob } p_2 \end{cases}$$



$$\mathbb{P}[\bar{X}=0] = \lambda(1-p_1) + (1-\lambda)(1-p_2)$$
$$= 1 - \bar{p}$$

$$\boxed{\bar{p} = \lambda p_1 + (1-\lambda) p_2}$$

$$\mathbb{P}[\bar{X}=1] = \lambda p_1 + (1-\lambda) p_2$$
$$= \bar{p}$$



$\lambda = \frac{1}{2}$

50% - 50% mixing

! $X \neq \mathcal{X}$ écrit → gras.

- SHANNON'S ENTROPY : $H(\mathbf{X}) = -\sum_{x \in X} p(x) \log_2 p(x)$

$$\underset{p(x)}{\boxed{X}} \longrightarrow x$$

- <u>JOINT ENTROPY</u> $(X,Y) \rightarrow (x,y)$ $\begin{matrix} x \in X \\ y \in Y \end{matrix}$ characterized by $p(x,y)$

$$H(X,Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2 p(X,y) \quad (\underline{\text{bit}})$$

- <u>CONDITIONAL ENTROPY</u> : $p(x,y) =$ joint proba

$p(y|x) = \dfrac{p(x,y)}{p(x)}$ cond proba

$\hookrightarrow H(Y|X) = \sum_{x \in X} p(x)\, H(Y|\cancel{X}=x)$

where $H(Y|X=x) = -\sum_{y \in Y} p(y|x) \log_2 p(y|x)$

DEF 1 $$H(Y|X) = -\sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2 p(y|x)$$

<u>CHAIN RULE</u> : $\boxed{H(X,Y) = H(X) + H(Y|X)}$

- DEMO : $H(X,Y) = -\sum_x \sum_y p(x,y) \log p(x,y)$

$$= -\sum_x \sum_y p(x,y) \log p(x) \cdot p(y|x)$$

$$= \underbrace{-\sum_x \underbrace{\sum_y p(x,y)}_{p(x)} \log p(x)}_{H(X)} \; + \; \underbrace{-\sum_x \sum_y p(x,y) \log p(y|x)}_{H(Y|X)}$$

- $p(x,y) = p(x) \cdot p(y|x)$

- $H(Y|X) = H(X,Y) - H(X)$ (Add uncertainty on Y once X is KNOWN)

DEF 2

CASE 1 : INDEPENDENT VARIABLES

$$p(x,y) = p(x)\cdot p(y) \longleftrightarrow p(y|x) = \frac{p(x,y)}{p(x)} = p(y)$$

$$\bullet\ H(Y|X) = -\sum_y \underbrace{\sum_x p(x,y)}_{p(y)} \log p(y) = H(Y)$$

$$\Rightarrow H(X,Y) = H(X) + H(Y)$$

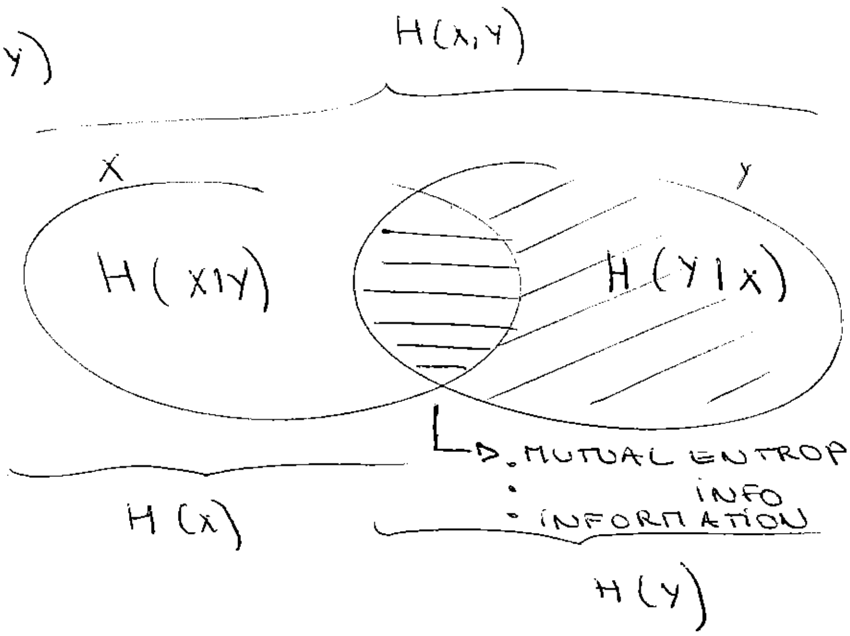
CASE 2 : FULLY CORRELATED VARIABLES $p(y|x) \begin{cases} = 1 & \text{for } y = y^* = f(x) \\ = 0 & y \neq y^* \end{cases}$

$$H(Y|X) = -\sum_x p(x) \underbrace{\sum_y p(y|x) \log p(y|x)}_{\underbrace{p(y^*|x)}_{=1} \log \underbrace{p(y^*|x)}_{=1} = 0}$$

$$= 0$$

Theorem : $H(Y|X) = 0$ iFF ($\longleftrightarrow$) Y is a function of X

$\bullet\ H(X,Y)$



$\bullet \begin{cases} H(X,Y) = H(X) + H(Y|X) \\ H(X,Y) = H(Y) + H(X|Y) \end{cases}$

$\Rightarrow$ • MUTUAL ENTROPY between X and Y
• INFO / INFORMATION between X and Y

$$\longrightarrow H(X) + H(Y|X) = H(Y) + H(X|Y)$$

INFO about X contained in Y $\leftarrow$

$$\underbrace{H(X) - H(X|Y)}_{\text{• Decrease of uncertainty on X due to learning Y}} = \underbrace{H(Y) - H(Y|X)}_{\text{INFO about Y cont. in X}} = H(X:Y) = I(X:Y)$$

DEF 1 MUTUAL ENT.

MUTUAL ENTROPY (INFORMATION)

• $X, Y \quad x \in X \quad y \in Y \quad p(X,y) \longrightarrow \begin{cases} p(x) = \sum_{y \in Y} p(x,y) \\ p(y) = \sum_{x \in X} p(x,y) \end{cases}$

DEF 2

$$\boxed{H(X:Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \frac{p(x) \cdot p(y)}{p(x,y)}}$$

Demo: • $p(y|x) = \dfrac{p(x,y)}{p(x)}$ et $p(x|y) = \dfrac{p(x,y)}{p(y)}$

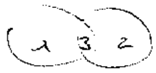
$$H(X:Y) = -\sum_x \sum_y p(X,y) \log \frac{p(x)}{p(x|y)}$$

$$= \underbrace{-\sum_x \underbrace{\sum_y p(x,y)}_{p(x)} \log p(x)}_{H(X)} + \underbrace{\sum_x \sum_y p(x,y) \log p(x|y)}_{-H(X|Y)}$$

= INFO ON X CONTAINED IN Y } ENTROPY SHARED BY X & Y
= " Y " X

• DEF 3 JOINT ENT

$$\boxed{H(X:Y) = \underset{1}{H(X)} + \underset{2}{H(Y)} - \underset{3}{H(X,Y)}}$$



• ENTROPY ≡ SELF-INFO

$$H(X:X) = H(X) - \underset{0}{\cancel{H(X|X)}} = H(X)$$

RELATIVE ENTROPY : (KULLBACK DISTANCE)

- 2 proba dist : $\begin{cases} p(x) \\ q(x) \end{cases}$ defined over X

$$D(p\|q) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)}$$

Measure "distance" between $p(x)$ and $q(x)$

D

- $\begin{cases} D(p(x) \| p(x)) = 0 \\ D(p(x) \| q(x)) \geq 0 \end{cases}$

FUNDAMENTAL INEQ.

- $D(p\|q) \neq D(q\|p)$
- Triangular inequality

- CONDITIONAL RELATIVE ENTROPY

$$D(p(y|x) \| q(y|x)) = \sum_x p(x) \sum p(y|x) \log \frac{p(y|x)}{q(y|x)}$$

- CHAIN RULE : $H(X,Y) = H(X) + H(Y|X)$

$$D(p(x,y) \| q(x,y)) = D(p(x) \| q(x)) + D(p(y|x) \| q(y|x))$$

↳ m̂ demo que la chaine rule precedente.

- INFORMATION INEQUALITY

- Let $\begin{cases} p(x) \\ q(x) \end{cases}$ be 2 proba distrib over X

- $D(p\|q) \geq 0$

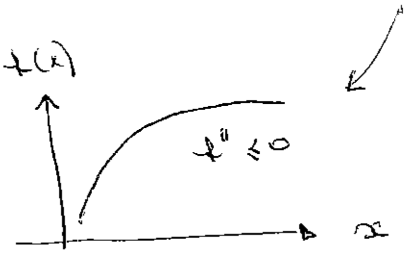
- $D(p\|q) = 0$ IFF $p(x) = q(x)$ $\forall x \in X$
$\longleftarrow$

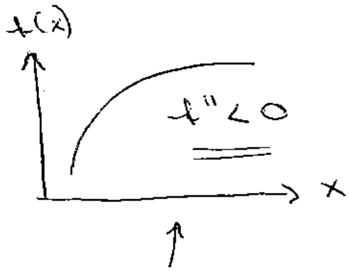
CODING THEORY 24.02

• JENSEN INEQUALITY

• IF f is CONCAVE FUNCTION then $E[f(x)] \leq f(E[X])$

where X = random variable.





• IF f is STRICTLY CONCAVE then $E[f(x)] = f(E[X])$

iff X = constant

$$-D(p\|q) = -\sum_{x\in X} p(x) \log \frac{p(x)}{q(x)}$$

• $A \equiv \{x : p(x) > 0\}$ = support of p

$B = \{x : q(x) > 0\}$ = support of q

$B \supset A$

LA $\quad = \sum_{x\in A} p(x) \log \frac{q(x)}{p(x)}$

$$= E\left[\log \frac{q(x)}{p(x)}\right]_{X \sim p(x)}$$

• log x is a strictly concave function of x $(x > 0)$

$$\leq \log\left(E\left[\frac{q(x)}{p(x)}\right]_{X\sim p(x)}\right) = \log\left(\sum_{x\in A} p(x) \frac{q(x)}{p(x)}\right)$$

devient $=$ si STRICTLY

• si $\frac{q(x)}{p(x)}$ = constant pour $\forall x$

$$= \log\Big(\underbrace{\sum_{x\in A} q(x)}_{\leq \sum_{x\in B} q(x) = 1}\Big) \leq 0 \implies D(p\|q) \geq 0$$

CONSEQ 1.

Distribution that MAXIMIZES entropy is UNIFORM DISTRIBUTIO

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

$$H(X) = \log_2 |\mathcal{X}| \quad \underset{\Longleftrightarrow}{\text{iff}} \quad p(x) = \frac{1}{|\mathcal{X}|}, \ \forall x \in \mathcal{X}$$

Demo: $\begin{cases} p(x) \\ u(x) = \frac{1}{|\mathcal{X}|} \end{cases}$

$$D(p\|u) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{u(x)} \geq 0$$

$$= \underbrace{\sum_{x \in \mathcal{X}} p(x) \log p(x)}_{-H(X)} + \underbrace{\sum_{x \in \mathcal{X}} p(x)}_{1} \log |\mathcal{X}| \geq 0$$

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

$$\& \ H(X) = \log |\mathcal{X}| \ \underline{\text{iff}} \ p(x) = u(x)$$

CONSEQ2: $(X;Y)$ pair of rand variables

- $H(X:Y) \geq 0$

$I = H(X:Y) = 0$ iff X and Y are independent variables.

Demo: $\begin{cases} H(X:Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in Y} p(x,y) \log \frac{p(x)\,p(y)}{p(x,y)} \\ D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \end{cases}$

$$I = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} = D\left(p(x,y) \,\|\, p(x) \cdot p(y)\right) \geq 0$$

$I = 0$ iff $p(x,y) = p(x) \cdot p(y) \iff X$ and Y are independent

Coding theory

3/3/2014

1. Asymptotic equipartition.

- $X_1, X_2, \ldots, X_n$ independent, identically distributed (i.i.d)
- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \simeq E(x)$ (esperance mathematic)
  (average) → for $n \gg 1$
- $\frac{1}{n} \log \frac{1}{P(x_1, \ldots, x_n)} = \frac{1}{n} \sum_{i=1}^{n} \log_2 \frac{1}{P(X_i)} \simeq E\left[\log \frac{1}{P(X_i)}\right] \equiv H(x)$ $(n \gg 1)$
  ↳ proba de réalisation

$P(x_1, \ldots, x_n) \simeq 2^{-n H(x)}$

- $X - [0, 1]$, $P(X=1) = P$, $P(X=0) = 1-P$
- $X^n = X_1, X_2, \ldots, X_n$
- $\prod_{i=1}^{n} P(X = X_i) = P^{\sum_i X_i} \cdot (1-P)^{n - \sum_{i=1}^{n} X_i}$
  $= P^4 (1-P)^2$
  $(1\,0\,1\,1\,0\,1)$
  $= P^{nP} (1-P)^{n(1-P)}$
  $= 2^{\log_2 P \cdot nP} \cdot 2^{\log_2 (1-P) \cdot n(1-P)}$
  $= 2^{-n[-P\log_2 P - (1-P)\log_2(1-P)]}$
  $= 2^{-n H(P, 1-P)}$
  ↳ entropy fond de la binaire (?)

- $\sum_i x_i \sim nP$
  $n - \sum_i x_i \sim n(1-P)$

- Ex: 10 11 00 11 000
  nP — $n(1-P)$

$N_{eT} = \binom{n}{nP} = \frac{n!}{(nP)!\,(n(1-P))!}$

- combien y a-t'il de séquences ≠ ?
- $n! \simeq n^n = 2^{n\log_2 n}$
- $= 2^{n\log_2 n - nP\log_2 nP - n(1-P)\log(1-P)}$
  $\simeq 2^{nH(P, 1-P)}$

• $P(X_1, \ldots, X_n) \cdot N_T \simeq 1 \quad n \gg 1$.

Asymptotic Equip. Theorem

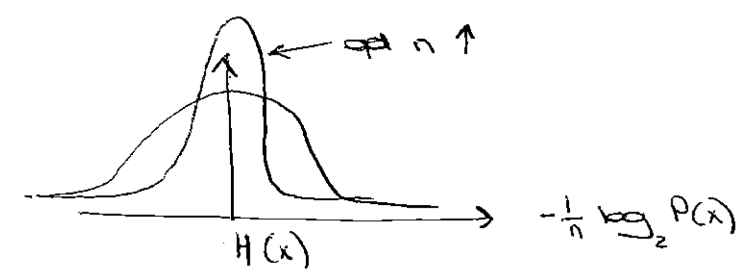
In $X_1, X_2, \ldots, X_n$ I.I.D

• $P(x) \Rightarrow \lim_{n \to \infty} \left[-\frac{1}{n} \log \frac{1}{P(X_1, X_2, \ldots, X_n)} \right] = H(x)$

Demo

$\rightarrow -\frac{1}{n} \log_2 \frac{1}{P(X_1, \ldots, X_n)} = -\frac{1}{n} \sum_{i=1}^{n} \log_2 \frac{1}{P(X_i)} \xrightarrow[n \to \infty]{LLN} E\left[-\log_2 P(x)\right] = H(x)$ □

L. Latch Number



---

2. Typical ensemble $A_\varepsilon^{(n)} \subset (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$

↳ param ↳ any sequences

$2^{-nH(x) - \varepsilon} \leq P(x_1, x_2, \ldots, x_n) \leq 2^{-nH(x) + \varepsilon}$

proba d'avoir une seq

• Prop 1 $H(x) - \varepsilon \leq -\frac{1}{n} \log_2 P(x_1, \ldots, x_n) \leq H(x) + \varepsilon$

if $(x_1, \ldots, x_n) \in A_\varepsilon^{(n)}$

• Prop 2 $P\left((x_1, x_2, \ldots, x_n) \in A_\varepsilon^{(n)}\right) > 1 - \varepsilon, \ n \gg 1$

• $\forall \delta > 0, \exists n_0 : \forall n \geq n_0$

$P\left(\left| -\frac{1}{n} \log \frac{1}{P(x_1, \ldots, x_n)} - H(x) \right| \leq \varepsilon \right) > 1 - \delta$

$\delta = \varepsilon$ $\quad P\left[(x_1, \ldots, x_n) \in A_\varepsilon^{(n)}\right] > 1 - \varepsilon$.

3/3/14

Coding theory

Prop 3 $|A_\varepsilon^{(n)}| \leq 2^{n(H(x)+\varepsilon)}$

$$1 = \sum_{x^n \in \mathcal{X}^n} P(x_1, \dots, x_n)$$

$$\geq \sum_{x^n \in A_\varepsilon^{(n)}} P(x_1, \dots, x_n)$$

$$\geq \sum_{x^n \in A_\varepsilon^{(x)}} 2^{-n(H(x)+\varepsilon)}$$

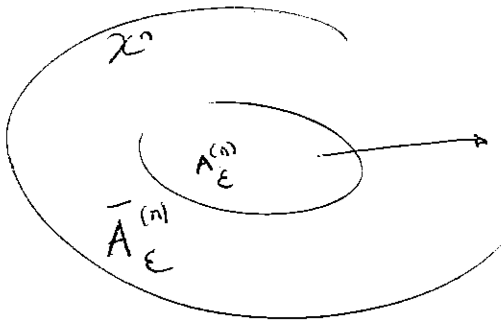
$$= |A_\varepsilon^{(n)}| \, 2^{-n(H(x)+\varepsilon)}$$

$$\Rightarrow |A_\varepsilon^{(n)}| \leq 2^{n(H(x)+\varepsilon)}$$

Prop 4 $|A_\varepsilon^{(n)}| > (1-\varepsilon)\, 2^{n(H(x)-\varepsilon)}$, $n >> 1$

$P\{A_\varepsilon^{(n)}\} > 1-\varepsilon$, $n >> 1$ $\Rightarrow$ $1-\varepsilon = \sum_{x^n \in A_\varepsilon^{(n)}} P(x^n) \leq \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(x)-\varepsilon)} =$

$$|A_\varepsilon^{(n)}| \, 2^{-n(H(x)-\varepsilon)}$$



la proba de se trouver dans cette ensemble $\simeq 1$

---

3. Data Compression

$\exists\, \varepsilon > 0$, $\exists n$, $\exists$ source code :

- $\forall\, x^n$ is represented by a seq of bits by a bijective map (invertible)
  (x^n ↑ seq)

- $E\left[\frac{1}{n} l(x^n)\right] < H(x) + \varepsilon$

Typical seq :
$\longrightarrow$ Prop 3 et 4 $\rightarrow$ $(1-\varepsilon)\, 2^{n(H-\varepsilon)} \leq |A_\varepsilon^{(n)}| \leq 2^{n(H+\varepsilon)}$

Flag si 1 $\rightarrow$ la seq $\in$ à l'ensemble $A_\varepsilon^{(n)}$
si 0 $\rightarrow$ $\notin$

Flag $\underbrace{\dots\dots}_{n(H+\varepsilon)+1}$ = size de la seq

Donc il faut $n(H+\varepsilon) + 2$ bits.

5 2

Non typical

Flag: $\underbrace{\cdots\cdots}_{n\log|\mathcal{X}|+1}$

$\underbrace{1\ \qquad\qquad\qquad}_{n\log|\mathcal{X}|+2}$

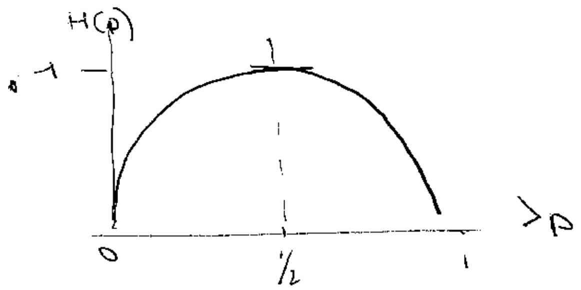
$$E(l(x^n)) = \sum_{x^n} P(x^n)\, l(x^n)$$

$$= \sum_{x^n \in A_\varepsilon^{(n)}} P(x^n)\, l(x^n) + \sum_{x^n \notin A_\varepsilon^{(n)}} P(x^n)\, l(x^n)$$

$$= \underbrace{\sum_{x^n \in A_\varepsilon^{(n)}}}_{\leq 1} P(x^n)\left[n(H+\varepsilon)+2\right] + \underbrace{\sum_{x^n \notin A_\varepsilon^{(n)}}}_{<\varepsilon} P(x^n)\left[n\log_2|\mathcal{X}|+2\right]$$

$$\leq n\Big(H + \underbrace{\varepsilon + \varepsilon\log|\mathcal{X}| + \tfrac{2}{n}}_{\varepsilon'}\Big)$$

- $H(X) << \log|\mathcal{X}|$
- si on est en binaire $\mathcal{X} = 2$ car 0 ou 1
- si on a des sequences random on ne peut pas les compresser... (?)



Coding theory

- Information inequality

- $D(p\|q) \geqslant 0$
  ↳ Kullback distance
- $D(p\|q) = 0$ iff $p = q$
- CONSEQ #1 : $H(X) \leqslant \log_2 |\mathcal{X}|$
  $=$ iff X is uniform over $\mathcal{X}$

- CONSEQ #2 : $I(X:Y) = D\big(p(x,y) \,\|\, p(x)p(y)\big) \geqslant 0$
  $= 0$ iff $p(x,y) = p(x)\,p(y)$
  $\Longleftrightarrow$ X and Y independant

$\Longleftrightarrow$ Subadditivity of entropy

$$\boxed{H(X,Y) \leqslant H(X) + H(Y)}$$

(= ssi X and Y are indep)

$$I(X:Y) = H(X) + H(Y) - H(X,Y) \geqslant 0$$

- CONSEQ #3 : $\underline{H(X|Y) \leqslant H(X)}$ The fact of learning y can only decrease entropy of X

$$I(X:Y) = H(X) - H(X|Y) \geqslant 0$$

- $H(X|Y) = \sum_y p(y)\, H(X|Y=y) \leqslant H(X)$

! $H(X|Y=y) \leqslant H(X),\ \forall y$ ⟶ WRONG

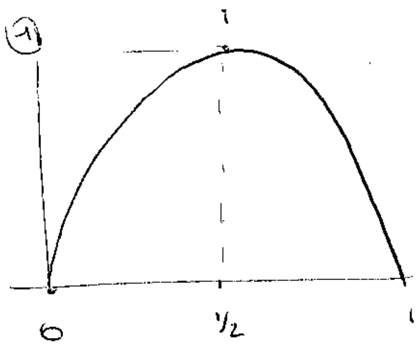
Example

p(x,y)		X = 1	X = 2	
Y	1	0	3/4	3/4 → (0 + 3/4)
	2	1/8	1/8	2/8 = 1/4
		1/8	7/8	

- $H(X) = \frac{1}{8}\log_2 8 + \frac{7}{8}\underbrace{\overset{\text{inverse}}{\log_2 \frac{8}{7}}}_{\log 8 - \log 7}$

$= 3 - \frac{7}{8}\log_2 7$

$= 0{,}544$ bits.



- $H(X|Y=1) = H_2[0,1] = 0 \quad < 0{,}544$

$H(X|Y=2) = H_2[1/2, 1/2] = 1/2 \quad \not< 0{,}544$

- $H(X|Y) = \underbrace{p(Y=1)}_{3/4} \times \underbrace{H(X|Y=1)}_{0} + \underbrace{p(Y=2)}_{1/4}\ \underbrace{H(X|Y=2)}_{1}$

$= 0{,}25$ bits $< 0{,}544$ bits.

- CONSEQ #4 : Concavity of H

- $H(\lambda p_1 + (1-\lambda) p_2) \geq \lambda H(p_1) + (1-\lambda) H(p_2)$

$p_1; p_2$ proba dist

$0 \leq \lambda \leq 1$

<u>Proof</u> :
- $X_1 \sim p_1$ ("distributed as")
- $X_2 \sim p_2$
- $\Theta \begin{cases} = 1 & \text{with proba } \lambda \\ = 2 & 1-\lambda \end{cases}$

$Z \equiv X_\Theta \sim \lambda p_1 + (1-\lambda) p_2$

- $H(Z) \geq H(Z|\Theta)$

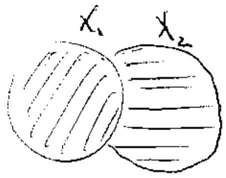
$$H(\lambda p_1 + (1-\lambda) p_2) \geq \underbrace{p(\Theta=1)}_{\lambda} \underbrace{H(Z|\Theta=1)}_{\underbrace{H(X_1)}_{=H(p_1)}} + \underbrace{p(\Theta=2)}_{1-\lambda} \underbrace{H(Z|\Theta=2)}_{\underbrace{H(X_2)}_{=H(p_2)}}$$

Entropy of MULTI-VARIABLE probability distributions

- collection of random variables $X_1 \dots X_n$

- CHAIN RULE for entropies : $H(X_1, \dots, X_n) = H(X_1) + H(X_2|X_1) + \dots$
$H(X_3|X_1, X_2) + \dots$
$H(X_n|X_1, \dots, X_{n-1})$

- $H(X_1, X_2) = H(X_1) + H(X_2|X_1)$



avoid double counting.

- $H(X_1, X_2, X_3) = H(X_1) + H(X_2, X_3 | X_1)$
$= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2)$

Donc $H(X_1, X_2, \dots X_n) = \sum_{i=1}^{n} H(X_i | X_1 \dots X_{i-1})$

$\left(! \text{ si } i=1 \quad H(X_1)\right)$

- SUBADDITIVITY of H

$H(X_1 X_2 \dots X_n) \leq H(X_1) + H(X_2) + \dots + H(X_n)$

Proof : $H(X_1 X_2 \dots X_n) = H(X_1) + \underbrace{H(X_2|X_1)}_{\leq H(X_2)} + \dots + \underbrace{H(X_n|X_1 \dots X_{n-1})}_{\leq H(X_n)}$

- $H(X_i) - H(X_i | X_1 \dots X_{i-1})$
$= I(X_i : X_1 \dots X_{i-1}) \geq 0$

X_i indep of $X_1 \dots X_{i-1}$

• CHAIN RULE for information

$$H(X_1 \dots X_n : Y) = H(X_1 : Y) + H(X_2 : Y \mid X_1) + \dots + H(X_n : Y \mid X_1 \dots X_{n-1})$$

Proof: $H(X_1 \dots X_n : Y) = H(X_1 \dots X_n) - H(X_1 \dots X_n \mid Y)$

$$\begin{cases} H(X_1 \dots X_n) = \sum_{i=1}^{n} H(X_i \mid X_1 \dots X_{i-1}) \\ H(X_1 \dots X_n \mid Y) = \sum_{i=1}^{n} H(X_i \mid X_1 \dots X_{i-1} Y) \end{cases}$$

$$= \sum_{i=1}^{n} \left\{ H(X_i \mid X_1 \dots X_{i-1}) - H(X_i \mid X_1 \dots X_{i-1} Y) \right\}$$

$$= \sum_{i=1}^{n} H(X_i : Y \mid X_1 \dots X_{i-1})$$

CONDITIONAL MUTUAL ENTROPY

$$H(X : Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z)$$

↑ MUTUAL ENT

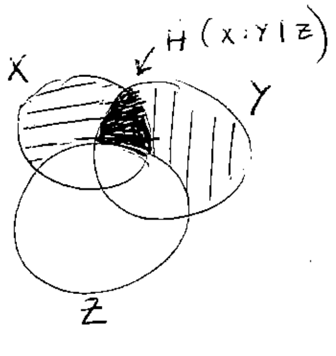
$$\left(\cdot \; H(X : Y_{|Z}) = H(X_{|Z}) - H(X \mid Y_{Z}) \right)$$

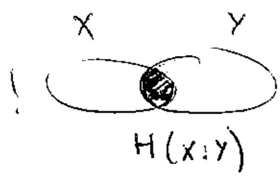
$$= H(X \mid Z) - H(XYZ) + H(YZ)$$

$$H(X \mid Z) \underbrace{- H(XYZ) + H(Z)}_{-H(XY \mid Z)} + H(Y \mid Z)$$

$$H(X : Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - H(X, Y \mid Z)$$

$$\left(H(X : Y) = H(X)_{|Z} + H(Y)_{|Z} - H(XY_{|Z}) \right)$$



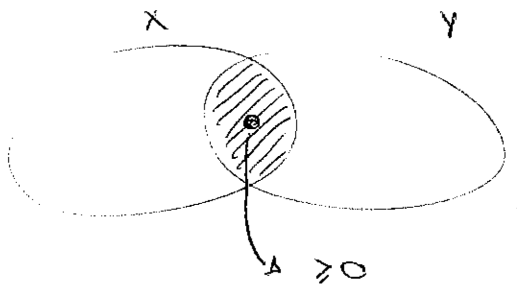


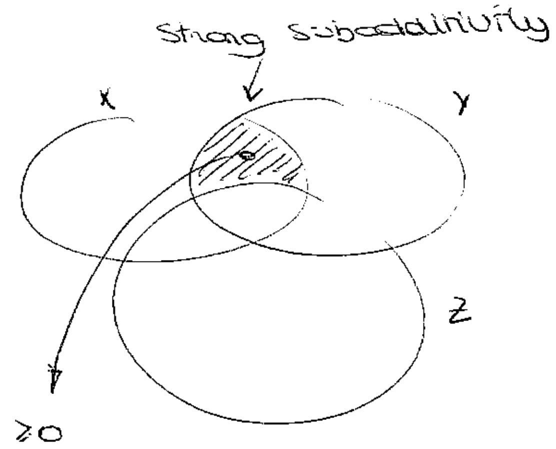
Coding theory

10.3.14

Summary

- $H(X) \longrightarrow H(X|Y)$
- $H(X:Y) \longrightarrow H(X:Y|Z)$
- $D(p\|q) \longrightarrow D(p(x|y)\|q(x|y))$
- Subadditivity $\longrightarrow$ Strong subadditivity





- $H(X:Y|Z) = H(X|Z) + H(Y|Z) - H(XY|Z)$

$$= -\sum_{xz} p(x,z) \log p(x|z) - \sum_{yz} p(y,z) \log p(y,z)$$

$\sum_y p(y|xz) = 1$ $\qquad$ $\sum_x p(x|yz) = 1$

$$+ \sum_{x,y,z} p(x,y,z) \log p(x,y|z)$$

$$= -\sum_{xyz} p(x,y,z) \log \frac{p(x|z)\,p(y|z)}{p(x,y|z)}$$

$$= \sum_z p(z) \underbrace{\sum_{x,y} p(x,y|z) \log \frac{p(x,y|z)}{p(x,z)\,p(y|z)}}_{\text{fixed } z}$$

$$= D\big(p(x,y|z) \,\|\, p(x|z)\,p(y|z)\big) \geq 0$$

- $I(X:Y) = D(p(x,y) \,\|\, p(x)p(y)) \geq 0$

$\iff H(XY) \leqslant H(X) + H(Y)$

$$\Big\downarrow$$

$\iff H(XY|Z) \leqslant H(X|Z) + H(Y|Z)$

10.3.14

Coding theory

- Chapter 3 : SOURCE CODING (DATA compression)

  ↳ Huffman code

- $\square \rightarrow a\,a\,z\,e$

  Morse : $E \rightarrow \cdot$ (E est la lettre la plus freq)

  $Q \rightarrow - - \cdot -$ (Q peu freq)

- Code C $\quad \mathcal{X} \rightarrow \mathcal{D}^*$

  $= \{a, b, c, \dots, z\}$ $\quad = \{0, 1, 2, \dots, D-1\}$ $\quad$ D = size of coding alpha

  $= \{0, 1\}$ when $D = 2$

- $C(x)$ = codeword associated with symbol x

  $l(x)$ = length of $C(x)$

  ↳ $C(a) = 00 \quad l(a) = 2$

  $C(b) = 01 \quad l(b) = 2$

  $C(c) = 111 \quad l(c) = 3$

Average length of code C for source X

$$L(C) = \sum_{x \in \mathcal{X}} p(x)\, l(x)$$

<u>Ex</u> : $\mathcal{X} = \{a, b, c, d\}$

$\mathcal{D} = \{0, 1\}$

- $\begin{cases} \mathbb{P}(X = a) = 1/2 & C(a) = 0 \\ \mathbb{P}(X = b) = 1/4 & C(b) = 10 \\ \mathbb{P}(X = c) = 1/8 & C(c) = 110 \\ \mathbb{P}(X = d) = 1/8 & C(d) = 111 \end{cases}$

$\rightarrow$ $L(C) = \frac{1}{2} \cdot \underset{\text{using 1bit}}{1} + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = \frac{7}{4}$ bits.

$H(X) = \dots = \frac{7}{4}$ bits. $\quad$ optimal code mais ne pourra jamais faire mieu que l'entropy.

CLASS of codes

① NON SINGULAR CODES

if $x_i \neq x_j \Rightarrow C(x_i) \neq C(x_j)$ les code words sont tous différents.

⊛ Extension C^* of code C

$$C^*(x_1, \dots, x_n) = C(x_1) \dots C(x_n)$$

Dans l'example ↑ $C^*(a\,d\,c) = C(a)C(c)C(d)$
$= 0110111$

② Uniquely decodable codes

whose extension C^* is non singular (le decoding doit être non ambigous)

③ Instantaneous codes

satisfy "prefix condition" !

- NO CODEWORD is PREFIX of another codeword
- ⇒ decoding can be performed step by step.

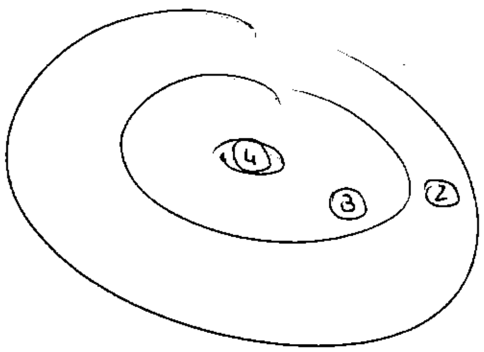
ex: $\underbrace{0}_{a}\underbrace{10}_{b}\underbrace{1111}_{d}\underbrace{010}_{b}\underbrace{10}_{b}$.

- ⇒ self punctuating code.

Coding theory

- Source coding

- Non singular $x_i \neq x_j \Rightarrow C(x_i) \neq C(x_j)$
- Uniquely decodable extension of C is non singular
- Instantaneous codes prefix condition.



1
2 Non sing
3 Uniq. Deco
4 Inst.

- Ex:

	1	2	3	4
X=1	0	0	10	0
X=2	0	010	00	10
X=3	1	01	11	110
X=4	1	10	110	111
		NON SING	Uniq Decod	Inst

→ ex: 10 | 110 | 0 on peut decoder INSTANTANEMENT. (pas d'ambiguités (11 avec 110)

→ on ne sait pas si c'est 11 ou 110

NON SING = tous les chiffres sont encod ≠

Uniq Decod → on peut decoder mais avec resolut° des ambiguités (prefix =)

KRAFT INEQUALITY : optimal code (shortest code word with Inst code)

Any instant. code with ~~et~~ coding alphabet of size D and with code word of length $l_1 \ldots l_m$ (m number)

SATISFIES $\boxed{\sum_{i=1}^{m} D^{-l_i} \leq 1}$

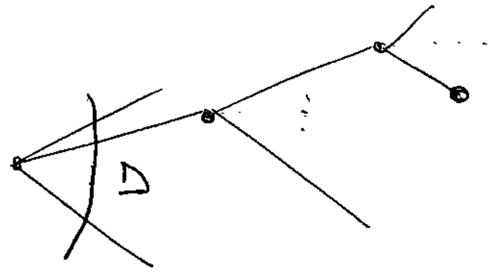
Necessary cond!

→ $= 1$ si on a le code OPTIMAL.

• Given $\{l_1 \dots l_m\}$ that satisfy $\sum_{i=1}^{m} D^{-l_i} \leq 1$, then there exists an instantaneous code with codewords of length $l_1 \dots l_m$ with coding alphabet of size D.

SUFFICIENT COND.

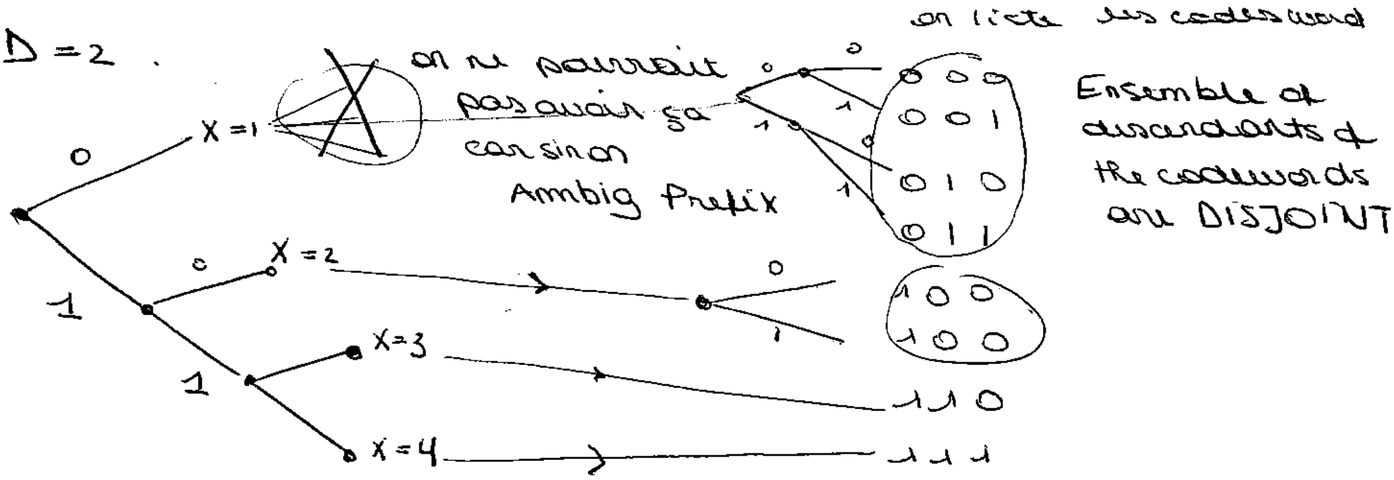
PROOF : "DECODING tree"



	④
X_1	0
X_2	10
X_3	110
X_4	111

tree of order D.

Ex : $D = 2$.



• l_{max} = length of the longest codewords ($l_{max} = 3$)

↳ codeword #i → $D^{l_{max} - l_i}$ descendants

$$\boxed{\sum_{i=1}^{m} D^{l_{max} - l_i} \leq D^{l_{max}}}$$

$$\Longleftrightarrow$$

$$\boxed{\sum_{i=1}^{m} D^{-l_i} \leq 1 .}$$

Coding theory.

Ex: (4)

x_1 0 $\rightarrow l_1 = 1$ $\quad l_{max} = 3$

x_2 10 $\quad l_2 = 2$

x_3 110 $\quad l_3 = 3$ $\quad \Rightarrow 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} \overset{?}{\leq} 1$

x_4 111 $\quad l_4 = 3$ $\quad \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = 1$

OPTIMAL.

LENGTH OF OPTIMAL INSTANTANEOUS CODE

- looking for $\{l_1 \dots l_m\}$ satisfying KRAFT ($\Rightarrow \exists$ inst code) such that average length L is minimum.

<u>MIN</u> $\quad L = \sum_{i=1}^{m} p_i l_i$ with l_i = integers

and constraint $\sum_{i=1}^{m} D^{-l_i} \leq 1$

Approx: $\begin{cases} l_i = \text{real} \\ \sum D^{-l_i} = 1 \end{cases}$

Lagrange multiplier method for constrained optimization.

$$J \equiv \sum_{i=1}^{m} p_i l_i + \lambda \left(\sum_{i=1}^{m} D^{-l_i} - 1 \right)$$

$$\frac{d}{dx} D^x = \frac{d}{dx} e^{\ln D x} = \ln D \cdot e^{\ln D x} = \ln D \cdot D^x$$

$$\begin{cases} \frac{\partial J}{\partial l_i} = p_i - \lambda \ln D \, D^{-l_i} = 0 \quad \forall i \\ \frac{\partial J}{\partial \lambda} = \sum_i D^{-l_i} - 1 = 0 \end{cases}$$

$$\Rightarrow D^{-l_i} = \frac{p_i}{\lambda \ln D} \rightarrow \sum_i \frac{p_i}{\lambda \ln D} = 1 \rightarrow \lambda = \frac{1}{\ln D} \rightarrow D^{-l_i} = p_i$$

$$\boxed{\bullet \; l_i^* = \underbrace{-\log_D p_i}_{\geq 0}} \neq \text{integer.}$$

- $L^* = \sum_i p_i l_i^* = -\sum_i p_i \log_D p_i$

$$\boxed{L^* = H_D(X)}$$

→ THEOREM :

The length L of any instantaneous code with coding alpha of size D for a source X must satisfy

$$\boxed{L \geq H_D(x)}$$

$L = H_D(x)$ iff source distribut° is "D-adic"

(p_i = integer power of $\frac{1}{D}$)

Proof : $L - H_D(x) = \sum_i p_i l_i + \sum_i p_i \log_D p_i$

$l_i = -\log_D(D^{-l_i})$

$= -\sum_i p_i \log_D D^{-l_i} + \sum_i p_i \log_D p_i$

$q_i = \frac{D^{-l_i}}{C}$ $\quad C = \sum_i D^{-l_i} \leq 1$

$= -\sum_i p_i \log q_i - \sum_i p_i \log C$
$+ \sum_i p_i \log_D p_i$

$= \underbrace{D(p \| q)}_{\geq 0}$ (relative entropy) $+ \underbrace{\log_D \frac{1}{C}}_{\geq 0}$

≥ 0 because $C \leq 1$.

- $L = H_D(x)$

iff $\begin{cases} D(p \| q) = 0 \\ C = 1 \end{cases}$ $\iff$ $p_i = D^{-l_i}$ (→ integer)

p_i is D adic distrib

ex : $D = 2$ $\quad \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\right)$

find optimal code?

① look for D adic distrib, "close to" p_i

② Check $l_i^* = -\log_D p_i$ = integer => KRAFT $\sum_i D^{-l_i^*} = \sum_i D^{\log_D p_i} = \sum_i p_i = 1.$

③ Build the code.

SHANNON'S CODE : $H(X) \leq L < H(X) + 1$

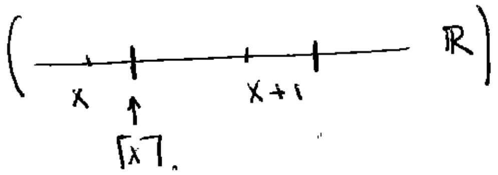
↳ penalty of 1 bit.

- $l_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$ (ceiling function of x, $\lceil x \rceil$ = smallest integer that is $\geq x$)
- KRAFT : $\sum_i D^{-l_i} = \sum_i D^{-\lceil \log_D \frac{1}{p_i} \rceil} \leq \sum_i D^{-\log_D \frac{1}{p_i}} = \sum_i p_i = 1.$

$$\begin{cases} \lceil x \rceil \geq x \\ -\lceil x \rceil \leq -x \\ D^{-\lceil x \rceil} \leq D^{-x} \end{cases}$$

$x \leq \lceil x \rceil < x + 1$

↳ Therefore there exists $\exists$ instaneous code $\equiv$ Shannon's code.



- $\log_D \frac{1}{p_i} \leq l_i < \log_D \frac{1}{p_i} + 1 \Rightarrow \sum_i p_i \log_D \frac{1}{p_i} \leq \sum_i p_i l_i < \sum_i p_i \left(\log_D \frac{1}{p_i} + 1\right)$

p_i. p_i. p_i.

$$\boxed{H_D(X) \leq L < H_D(X) + 1}$$

↳ The worst case penalty is 1 bit. (car on n'est pas D-adic, si c'était le cas on atteindrait l'entropie $H_D(X)$)

Theorem Length L^* of ~~For~~ optimal code with codewords length $l_1^*, \dots, l_m^*$ and coding alpha of size D MUST SATISFY

$$\boxed{H_D(X) \le \underbrace{L^*}_{\text{length of optimal code}} < H_D(X) + 1}$$

- $H_D(X) \le L^*$
- $L^* \le L_{Sh}$

$$\left.\begin{matrix} H_D(X) \le L^* \\ L^* \le L_{Sh} \end{matrix}\right\} \quad \underline{H_D(X) \le L^* < L_{Sh} < H(X) + 1}$$

Reduction de la pénalité

BLOCK CODING

$\boxed{p_i} \longrightarrow \overbrace{x_1 \; x_2 \dots x_n}^{\text{bloc (sequence)}}$

iid
indep identically distrib.

bloc $\in \mathcal{X}^n$

L_Δ shannon's code.

LOWER BOUND

Upper Bound

$$\underbrace{\frac{H(X_1 \dots X_n)}{n}}_{H(X)} \le \underbrace{\frac{1}{n} \sum p(x_1, \dots x_n)\, l(x_1, \dots x_n)}_{L_n = \text{mean code length per source symbol.}} < \underbrace{\frac{H(X_1 \dots X_n) + 1}{n}}_{H(X) + \frac{1}{n}}$$

average of bits I need to encode n bits.

$$\left(\begin{aligned} H(X_1 \dots X_n) &= H(X_1) + \dots + H(X_n) \\ &= n\, H(X) \end{aligned} \right)$$

$\hookrightarrow$ var indep

Coding Theory

24.03.

SHANNON'S CODE (sub optimal)
HUFFMAN'S CODE (optimal)

$$L \Rightarrow l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil \qquad H(X) \leq \underbrace{E[l(x)]}_{L} < H(X) + 1 \text{ bit}$$

Block Code $\quad \underset{\text{iid}}{\boxed{X}} \rightarrow$ MESSAGES

0110 0 1101

010 0 111

- $L_n = \frac{1}{n} E[l(x_1, x_2, \ldots, x_n)]$ = mean length per source symbol

- $\underbrace{H(X_1 \ldots X_n)}_{} \leq \underbrace{E[l(X_1 \ldots X_n)]}_{n L_n} < \underbrace{H(X_1 \ldots X_n)}_{n H(X)} + 1$

$= H(X_1) + \ldots + H(X_n)$
$= n H(X)$

$$\Rightarrow \boxed{H(X) \leq L_n < H(X) + \underbrace{\frac{1}{n}}_{\rightarrow 0 \text{ (tend vers 0)}}}$$

Shannon's first Theorem.

$$H(X) \leq L_n^{OPT} \leq L_n^{SHANNON} < H(X) + \frac{1}{n}$$

- $\frac{H(X_1 \ldots X_n)}{n} \leq L_n < \frac{H(X_1 \ldots X_n)}{n} + \frac{1}{n}$

Process is STATIONARY $\quad \lim_{n \rightarrow \infty} \frac{H(X_1 \ldots X_n)}{n} = H(\mathcal{X}) \equiv$ ENTROPIC RATE OF SOURCE

$= H(X_i \mid X_{i-1})$

MARKOV

$$\boxed{H(\mathcal{X}) \leq L_n < H(\mathcal{X}) + \frac{1}{n}}$$

Coding a source of UNKNOWN DISTRIBUTION

$\boxed{X} \longrightarrow p(x)$ = actual prob. distrib (unknown)

$q(x)$ = best estimate of $p(x)$ used for encoding

$$l(x) = \left\lceil \log \frac{1}{q(x)} \right\rceil \quad \text{SHANNON'S CODE}$$

$$\boxed{a \leqslant \lceil a \rceil < a+1}$$

- $$E[l(x)]_p = \sum_x p(x)\, l(x)$$
$$= \sum_x p(x) \left\lceil \log \frac{1}{q(x)} \right\rceil \quad \text{(because we have just an approximation)}$$
$$\geqslant \sum_x p(x) \log \frac{1}{q(x)} \cdot \frac{p(x)}{p(x)}$$
$$\geqslant \underbrace{\sum_x p(x) \log \frac{p(x)}{q(x)}}_{D(p\|q) \geqslant 0} + \underbrace{\sum_x p(x) \log \frac{1}{p(x)}}_{H(p)}$$

$D(p\|q)$ = penalty

- Si l'estimation est loin de $p(x)$, penalty ↑

- $$E[l(x)]_q \leqslant \underbrace{\sum_x p(x) \log \frac{1}{q(x)} \frac{p(x)}{p(x)}}_{D(p\|q)+H(p)} + \underbrace{\sum_x p(x)}_{1 \text{ bit}}$$

$$L_D \; \boxed{H(p) + \underbrace{D(p\|q)}_{\geqslant 0} \leqslant L^{SH} < H(p) + \underbrace{D(p\|q)}_{\geqslant 0} + \frac{1}{n}}$$

Universal CODE (adaptive code)

$D(p\|q) \longrightarrow 0$

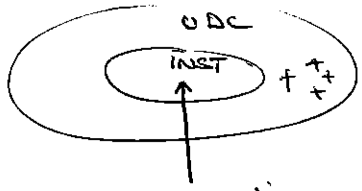
Ex : LEMPEL-ZIV CODE

Coding Theory

MAC MILLAN'S THEOREM

- ANY UDC satisfies

Kraft $\sum_i D^{-l_i} \leq 1$



$\sum_i D^{-l_i} \leq 1$

$H(X) \leq L^{\text{instantaneous}}$

HUFFMAN'S CODE (OPTIMAL SOURCE CODE)

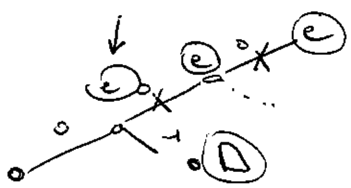
$\mathcal{X} = \{a, b, c, d, e\}$

$D = 2$

X	$\mathbb{P}(X = x)$		
a	0,25		
b	0,25		
c	0,20		
d	0,16	... 0	d/e 0,30
e	0,14	... 1	

on les lie entre elles et forme 1 entité !

NOT bcs the same prefix with D.



NOT possible si il y avait childs.

↳ Il faut que sa taille soit la m̂

(2) In the optimal instaneous code, the 2 least frequent symbols are encoded into the 2 longest codewords having the same lenght

(3) WNLG, 2 longest codewords share same prefix and differ by the last bit

(1) voir p3.

$\Rightarrow$

X	$\mathbb{P}(X = x)$		
d/e	0,30		
a	0,25		
b	0,25	... 0	b/c 0,45
c	0,20	... 1	

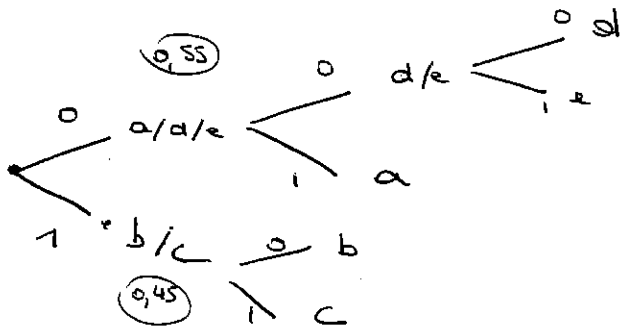
$\Rightarrow$

b/c	0,45		a/d/e
d/e.	0,30	... 0	0,55
a	0,25	... 1	

⇓

a d e	0,55	$\longrightarrow 0$	1.
b/c	0,45	$\longrightarrow 1$	

- on go backward



X	$C(x)$	$l(x)$
a	01	2
b	10	2
c	11	2
d	000	3
e	001	3

Kraft : $\underbrace{2^{-2}+2^{-2}+2^{-2}+2^{-3}+2^{-3}}_{1} \leq 1$

- OPTIMAL CODE $\quad H(X) \leq L^* < H(X)+1$

$$L=\sum_x p(x)\, l(x)$$

$$= \underbrace{(0{,}25+0{,}25+0{,}20)}_{0{,}70}\,2 + \underbrace{(0{,}16+0{,}14)}_{0{,}30}\,3$$

$= 2{,}30$ bits/symbol $\longrightarrow$ excess of $0{,}6\,\%$ $\rightarrow$ GOOD.

$$H(X) = -\sum_x p(x)\log p(x) = 2.2855 \text{ bits/symbol}$$

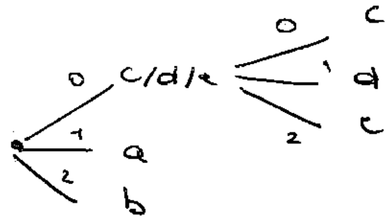
Example $D=3 \quad \{0,1,2\}$.

$\hookrightarrow$ au lieu de prendre les 2 derniers on prend les 3 derniers...

X	$\mathbb{P}(X=x)$	$l(x)$	
a	0,25	2	
b	0,25	2	
c	0,20	2	... 0
d	0,16	3	... 1
e	0,14	3	... 2

$\}$ c/d/e 0,50 $\Rightarrow$

c/d/e	0,5	... 0	
a	0,25	... 1	$\}$ 1
b	0,25	... 2	



X	c(x)	l(x)
a	1	1
b	2	1
c	00	2
d	01	2
e	02	3

- $L = 0{,}50 \times 1 + 0{,}50 \times 2$
  $= 1.50$ trits / symbol   ternary digits.

- $H(X) = -\sum_x p(x) \log_3 p(x) = 1{,}442$ trits/symb.

OPTIMALITY $L(C) \leq L(\text{any } C')$

- Lemma : For any source distribution $\{p_1, \dots, p_m\}$
  $\exists$ instantaneous optimal code C_m
  satisfying the following properties.

(1) If $p_j > p_k$, then $l_j \leq l_k$

$\overset{*}{\Rightarrow}$ (1) OPT code $C_m \to C_m'$ by interchanging $C(j) \leftrightarrow C(k)$

$$\underbrace{L(C_m') - L(C_m)}_{\geq 0} = \sum_i p_i l_i' - \sum_i p_i l_i$$

$$= p_j l_j' + p_k l_k' - p_j l_j - p_k l_k$$

$$= p_j l_k + p_k l_j - p_j l_j - p_k l_k$$

$$= \underbrace{(p_j - p_k)}_{>0} \Rightarrow \underbrace{(l_k - l_j)}_{\geq 0}$$

if $p_j > p_k \Rightarrow L_j \geq L_k$

P	C_{m-1}		C_m	
$p_1' = p_1$	w_i'	l_i'	$w_1 = w_1'$	$l_1 = l_1'$
$\vdots$	$\vdots$		$\vdots$	
$p_{m-2}' = p_{m-2}$	w_{n-2}'	l_{n-2}'	$w_{n-2} = w_{n-2}'$	$l_{n-2} = l_{n-2}'$
$p_{m-1}' = p_{m-1} + p_m$	w_{n-1}'	l_{n-1}'	$w_{n-1} = w_{n-1}'0$	$l_{n-1} = l_{n-1}' + 1$
			$w_n = w_{n-1}'1$	$l_n = l_{n-1}' + 1$

$$L(C_m) = \sum_{i=1}^{m} p_i l_i = \sum_{i=1}^{m-2} p_i' l_i' + p_{m-1} l_{m-1} + p_m l_m$$

$$+ \underbrace{(p_{m-1} + p_m)}_{p_{m-1}'} (l_{m-1}' + 1)$$

$$= \sum_{i=1}^{m-1} p_i' l_i' + p_{m-1} + p_m$$

$$= L(C_{m-1}) + \underbrace{p_{m-1} + p_m}_{\text{constant}}$$

Coding Theory

Shannon's CODE sub optimal $H(X) \leq L^{SH} < H(X)+1$

HUFFMAN's CODE optimal $H(X) \leq L^{*} < H(X)+1$

- $L^{*} \leq L^{SH}$
- Source $\{P_1, \dots, \underbrace{P_{m-1}, P_m}\}$ $L(C_m)$ $\Rightarrow L(C_m) = L(C_{m+1}) + \underbrace{P_{m-1} + P_m}_{est}$

Source $\{P_1', \dots, P_{m-2}', \underbrace{P_{m-1} + P_m}_{P_{m-1}'}\}$ $L(C_{m-1})$

COMPARAISON SHANNON vs HUFFMAN

Ex 1.

			Shannon		Huffman
$a \to 0{,}9999$	$l_a = \lceil \log_2 \overbrace{\frac{1}{P_a}}^{1,\dots} \rceil =$		$\boxed{1}$	$>$	$\boxed{1}$
$b \to 0{,}0001$	$l_b = \lceil \log_2 \underbrace{\frac{1}{P_b}}_{10\,000} \rceil =$		$\boxed{14}$	$\geq$	$\boxed{1}$

$\begin{cases} a \Rightarrow 0 \\ b \Rightarrow 1 \end{cases}$

Ex 2

Shannon

a $1/3$ ⟶ $1/3$

b $1/3$ ⟶ $1/3$ } 0

c $1/4$ } 0 $\Rightarrow c/d = 1/3 \to 1/3$ } 1 · b/c/d $2/3$

d $1/12$ } 1

bcd $2/3$ } 0

a $1/3$ } 1

codewords.

a	1
b	00
c	010
d	011

$\Rightarrow$

huffman

a	$1/3$	$l_a = \lceil \log_2 3 \rceil = 2$
b	$1/3$	$l_b = 2$
c	$1/4$	$l_c = \lceil \log_2 4 \rceil = 2$
d	$1/12$	$l_d = \lceil \log_2 12 \rceil = 4$

	shannon	huffman
a	1	2
b	2	2
c	3	2
d	3	4

$\Rightarrow$ contre exemple où Shannon $\ngtr$ huffman

$L^{Huff} = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{4} \cdot 3 + \frac{1}{12} \cdot 3 = 2$ bits

$L^{Shann} = 2 + 2 \cdot \frac{1}{12} = 2 + \frac{1}{6}$ bits

⇒ Ex 2 : other way.

a	1/3		
b	1/3		
c	1/4	0	cd 1/3
d	1/12	1	

$\left.\begin{matrix}0\\1\end{matrix}\right\}$ ab 2/3, cd 1/3 $\left.\right\}$ 0, 1

	huff
a	0 0
b	0 1
c	1 0
d	1 1

CHAPTER : CAPACITY of NOISY CHANNELS

DISCRETE

$W \longrightarrow$ [ENCODER] $\xrightarrow{X^n}$ [NOISY CHANNEL] $\xrightarrow{Y^n}$ [DECODER] $\xrightarrow{\hat{W}}$

rand. variab

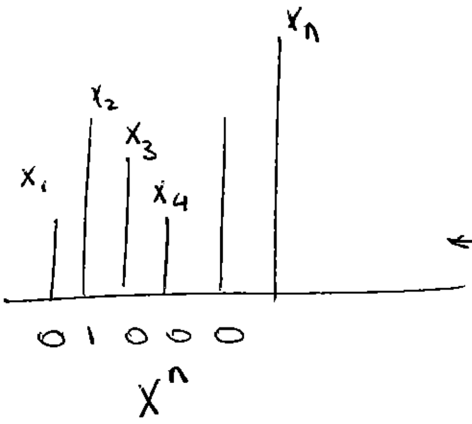
$X^n \equiv X_1 \dots X_n$

$p(y|x)$

$Y^n = Y_1 \dots Y_n$

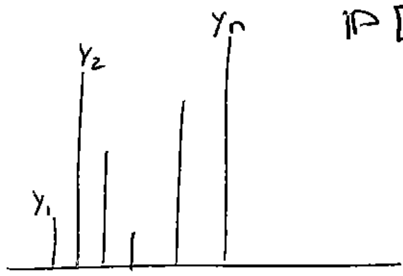
• estimator

$P_{error} = \mathbb{P}[\hat{W} \neq W]$



$p(y|x)$

← bit stream ⇒



• Restrictions
i) Discrete channels
ii) Memoryless channels ($x^n = x_1 \dots x_n$)

$P(y^n | x^n)$ what's the output knowing the input.

$$= \prod_{i=1}^{n} p(y_i | x_i)$$

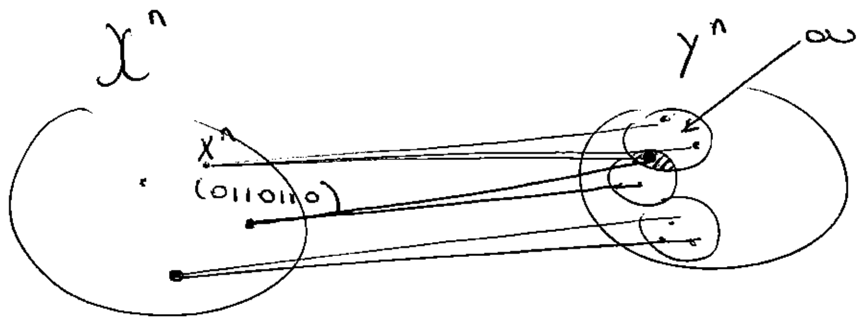
• Highest number of message "DISTINGUISHABLE" $\leq 2^{n \cdot C}$

channel capacity.

$$\log_2 (n\pi) \leq n \cdot C$$

$$\frac{1}{n} \log (n\pi) \leq C$$ → INFO SENT PER USE OF CHANNEL

Coding theory



we cannot disting. anymore x^n

→ To AVOID but very hard so we have to reduce the probab. to have overlapping.

- 4 msg

$$\left| \begin{array}{lcl} 00 & \longrightarrow & 0010 \\ 01 & \longrightarrow & 0100 \\ 10 & \longrightarrow & 1111 \\ 11 & \longrightarrow & 1010 \end{array} \right.$$

$\underbrace{\quad}_{2}$ $\qquad$ $\underbrace{\quad}_{4}$

2 logical bits
4 physical bits. → $\underset{R}{\text{Rate}} = \frac{1}{2} \leq C$

I) DEFINITIONS

Discrete memoryless channel

- input alphabet X
- output Y
- matrix of transition probabilities . P
  $p(y|x)$

- $P(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$

① Informational Capacity $C \equiv \max_{p(x)} I(X:Y)$

$x \xrightarrow{\boxed{p(y|x)}} y$

[shannon]

② Operational Capacity ≡ highest trans. rate R (bits/used channel) that is achievable with arbitrarily small error probability. P_e

Shannon's 2^{nd} theorem:

$$\boxed{C_{info} = C_{operat.}}$$

Ex 1 : NOISELESS BINARY CHANNEL.

bits

$0 \rightarrow 0$ $\qquad$ $C_{oper} = 1$ bit / use of channel

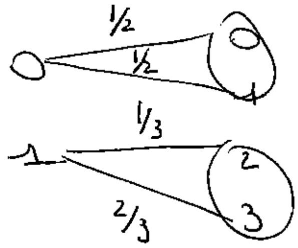
$1 \rightarrow 1$

① $I(X:Y) = H(X) - \underbrace{H(X|Y)}_{0}$

$C_{info} = \max\limits_{p(x)} H(x) = \log_2 2 = 1$ bit / use of channel.

$p^*(x) = \{1/2, 1/2\}$. → proba d'avoir un bit 1 ou 0 sur le channel.

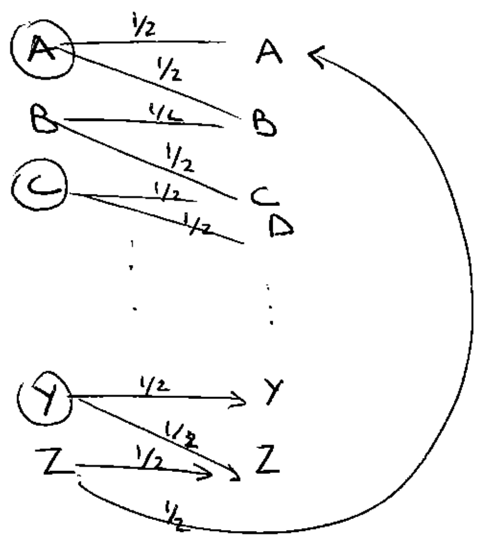
Ex 2 : NOISY CHANNEL WITHOUT OVERLAP



- $C_{oper} = 1$ bit / use of channel
- $I(X:Y) = H(x) - \underbrace{H(X|Y)}_{0}$
- $C_{info} = \max\limits_{p(x)} H(x) = 1$ bit / use of channel

$p^*(x) = \{1/2, 1/2\}$.

Ex 3 : NOISY CHANNEL WITH OVERLAP $\qquad$ ("noisy typewriter")



- $C_{oper} = \log_2 13$ bits / use of channel

$I(X:Y) = H(Y) - H(Y|X)$

$= H(Y) - \underbrace{\sum_{x \in \mathcal{X}} p(x)}_{1} \underbrace{H(Y|X=x)}_{1\ \forall x}$

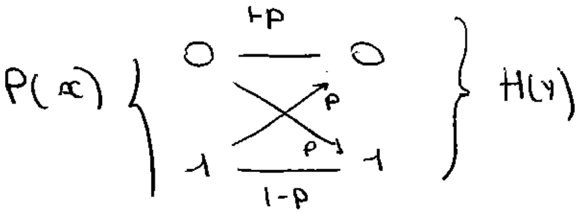
$= H(Y) - 1$

- $C_{info} = \max\limits_{p(x)} \underbrace{H(Y) - 1}_{\leq \log_2 26}$ $\qquad$ $H(Y) \leq \log|\mathcal{X}|$

$= \log_2 26 - 1 = \log_2 \frac{26}{2} = \log_2 13$ bits/use

$p^*(x) = \{1/26, \dots, 1/26\}$.

Ex 4 Binary Symmetric Channel 31/3.



p = bit flip probability

$0 \leq p \leq 1$

$$H(X:Y) = H(Y) - H(Y|X)$$
$$= H(Y) - \sum_{x=0}^{1} \underbrace{P(x)}_{1} \underbrace{H(Y|X=x)}_{H_2(p)\ \forall x}$$
$$= H(Y) - H_2(p)$$

• $H_2(p) \equiv H_2[p, 1-p]$
$$= -p \log p - (1-p)\log(1-p)$$

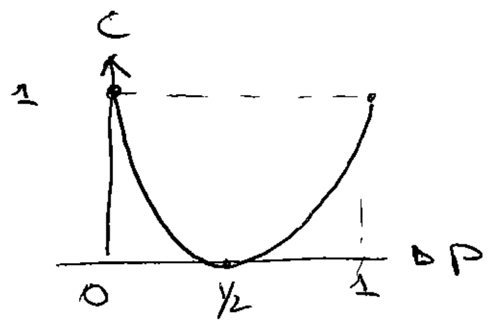
• $\mathbb{P}(y=0) = \underbrace{\mathbb{P}(x=0)}_{1/2} \cdot (1-p) + \underbrace{\mathbb{P}(X=1)}_{1/2}\, p = 1/2$

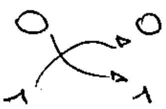
• $C = \max\limits_{p(x)} H(Y) - H_2(p) \leq 1 - H_2(p)$

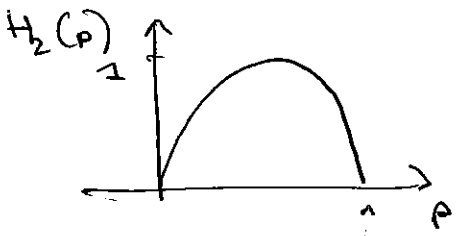
$\leq \log_2 |y| = 1$

$C = 1 - H_2(p)$ X equiprobl = Y equiprob.

$p^*(x) = \{1/2, 1/2\}$.





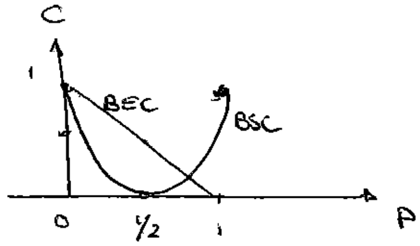


$H(Y) \leq \log(y)$

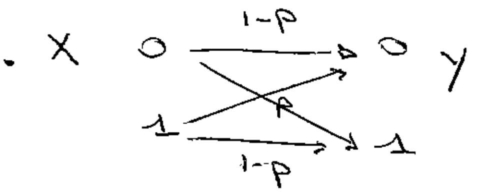
CHANNEL CAPACITY Code theory 28.04.

- **Binary symmetric channel**

p = bit flip (error) proba
$0 \le p \le 1$



$$C = \max_{p(x)} I(X:Y)$$



$$C_{BSC} = 1 - H_2(p) \qquad p^*(x) = \tfrac{1}{2} \quad (x = 0,1)$$

$$H_2(p) \equiv -p \log_2 p - (1-p)\log_2(1-p)$$

- **Binary erasure channel**

- α = erasure prob
  $0 \le \alpha < 1$

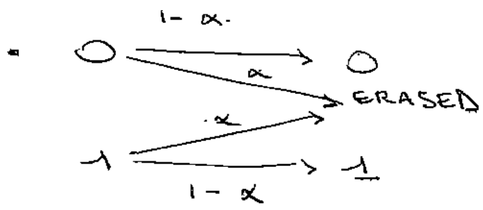
$$\underbrace{I(X:Y)}_{\substack{\max\\ p(x)\\ \uparrow \\ \max \pi}} = H(y) - H(Y|X)$$

$$= H(y) - \sum_{x=0}^{1} p(x)\, \underbrace{H(Y|X=x)}_{-\alpha\log\alpha - (1-\alpha)\log(1-\alpha)}$$

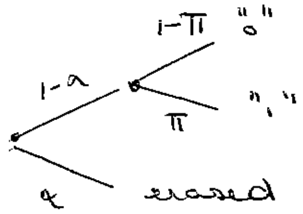
$$= H(Y) - H_2(\alpha) \qquad \equiv H_2(\alpha) \quad \forall \alpha$$

$$\le \log|Y| - H_2(\alpha)$$

$$\le \log_2 3 - H_2(\alpha)$$



$$\begin{cases} \mathbb{P}(X=0) = 1-\pi \\ \mathbb{P}(X=1)\ \pi \end{cases}$$



$$\to \begin{cases} \mathbb{P}(Y=0) = (1-\pi)(1-\alpha) \\ \mathbb{P}(Y=1) = \pi\ (1-\alpha) \\ \mathbb{P}(Y=\text{erased}) = (1-\pi)\alpha + \pi\alpha = \alpha \end{cases}$$

"grouping RULE"

$$\Rightarrow H(Y) = H_3\Big[\underbrace{(1-\pi)(1-\alpha)}_{0}, \underbrace{\pi(1-\alpha)}_{1}, \underbrace{\alpha}_{\text{erasure}}\Big] = \begin{matrix} H_2(\alpha) + (1-\alpha).H_2(\pi) \\ + \alpha \cancel{\log 1}_{0} \end{matrix}$$

$$\Rightarrow I(X:Y) = (1-\alpha)\, H_2(\pi)$$

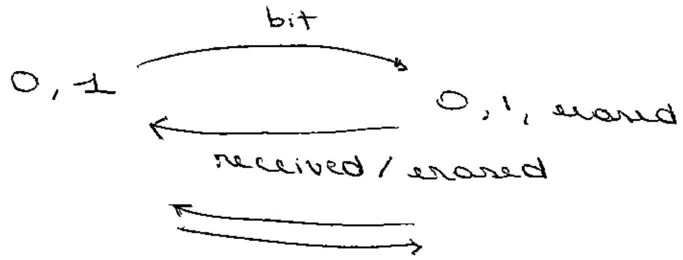
$$\Rightarrow C = \max_{\pi} (1-\alpha)\, H_2(\pi)$$

$$= 1-\alpha; \quad p^*(x) = \frac{1}{2} \quad (x=0,1)$$

$(1-\alpha = \text{prob of no erasure})$

• BIN ERASURE CHANNEL WITH FEEDBACK

↳ feedback without error.



$$\text{Nr of uses of channel} = (1-\alpha).1 + \alpha(1-\alpha).2 + \alpha^2(1-\alpha).3 + \dots$$

$$= (1-\alpha) \cdot \sum_{k=1}^{\infty} k\, \alpha^{k-1}$$

$$= (1-\alpha) \frac{\partial}{\partial \alpha} \left(\sum_{k=1}^{\infty} \alpha^k \right)$$

$$= (1-\alpha) \frac{\partial}{\partial \alpha} \left(\frac{\alpha}{1-\alpha} \right)$$

$$= (1-\alpha) \cdot \frac{(1-\alpha) + \alpha}{(1-\alpha)^2}$$

$$= \frac{1}{1-\alpha}$$

- $\sum_{k=0} \alpha^k = \frac{1}{1-\alpha}$
- $\frac{1}{1-\alpha} - 1 = \frac{1-(1-\alpha)}{1-\alpha} = \frac{\alpha}{1-\alpha}$
- $\frac{1}{1-\alpha}$ uses of channel to transmit 1 bit
- 1 uses of channel to transmit $(1-\alpha)$ bits

$C_{\text{without feedback}} = 1-\alpha$

$C_{\text{feedback}} = 1-\alpha$

THEOREM Feedback does not enhance the capacity of memoryless channel.

$$P(y^n | x^n) = \prod_{i=1}^{n} p(y_i | x_i)$$

Proprieties of C (arbitrary channel)

① $C \geq 0$ $\quad I(X:Y) \geq 0 \longleftrightarrow H(X,Y) \leq H(X) + H(Y);\ I(X:Y) = H(X) + H(Y) - H(X,Y)$

② $C \leq \log_2 |X|$ $\qquad I(X:Y) = H(X) - \underbrace{H(X|Y)}_{\geq 0} \leq H(X) \leq \log_2 |X|$

③ $C \leq \log_2 |Y|$ $\qquad I(X:Y) = H(Y) - \underbrace{H(Y|X)}_{\geq 0} \leq H(Y) \leq \log_2 |Y|$

④ $I(X:Y)$ is a continuous, concave function of $p(x)$

$\rightarrow p(x) = \lambda p_1(x) + (1-\lambda) p_2(x) \quad 0 < \lambda < 1$

$\quad \hookrightarrow I \geq \lambda I_1 + (1-\lambda) I_2$

$\rightarrow$ numerical optimization techniques.

- SYMMETRIC CHANNELS $\begin{cases} \mathcal{X} \\ Y \\ p(y|x) = \text{matrix of transition probabities.} \end{cases}$

$- \; X \rightarrow (\cdot \cdot \cdot \cdot \cdot)$, $\uparrow y$

channel SYMMETRIC

IF $\begin{cases} \text{all rows of } p(y|x) \text{ are permutations of each other.} \\ \text{" columns ——— " ———} \end{cases}$

Ex.: $|\mathcal{X}| = |Y| = 3$

$$P(y|x) = \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}$$

channel weakly symmetric

IF $\begin{cases} \text{all rows } p(y|x) \text{ are permutations of each other} \\ \text{all columns sum To SAME NUMBER} \end{cases}$

$$\sum_x p(y|x) = \text{constant} \quad \forall y.$$

Ex: $|\mathcal{X}| = 2, \; |y| = 3$

$$p(y|x) = \begin{pmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{pmatrix}$$
$$\quad 2/3 \quad 2/3 \quad 2/3$$

THEOREM A (weakly) symmetric channel has CAPACITY $C = \log_2 |Y| - H_2(\pi)$ WHICH is reached for Uniform Input Dist. $p^*(x)$

DEMO

$$p(y|x) = \begin{pmatrix} \overset{\pi}{\overline{\cdots\cdots}} \\ \cdots\cdots \\ \cdots\cdots \\ \cdots\cdots \end{pmatrix} x$$

$$I(X:Y) = H(Y) - H(Y|X)$$
$$= H(Y) - \sum_{x \in \mathcal{X}} p(x) \underbrace{H(Y|X=x)}_{H_2(\pi) \; \forall x}$$
$$= H(Y) - H_2(\pi)$$
$$\leq \log |Y| - H_2(\pi)$$

$$p^*(x) = \frac{1}{|\mathcal{X}|} \quad \text{uniform distrib}$$

$$\rightarrow p(y) = \sum_x p(y|x)\, p^*(x)$$

$$= \frac{1}{|\mathcal{X}|} \cdot \underbrace{\sum_x p(y|x)}_{\text{cont}} \quad \forall y \qquad \rightarrow C = \log_2 |\mathcal{Y}| - H_2(\pi)$$

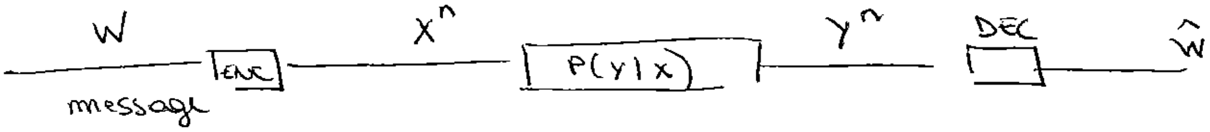
$$= \frac{1}{|\mathcal{Y}|} \quad \text{uniform}$$

Ex 1: $C = \log_2 3 - H[0.3, 0.2, 0.5] \simeq 0.1$ bits/use

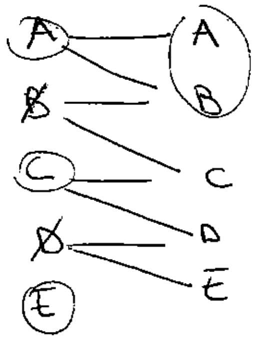
Ex 2: $C = \log 3 - H[1/3, 1/6, 1/2] \simeq 0.13$ bits/use.

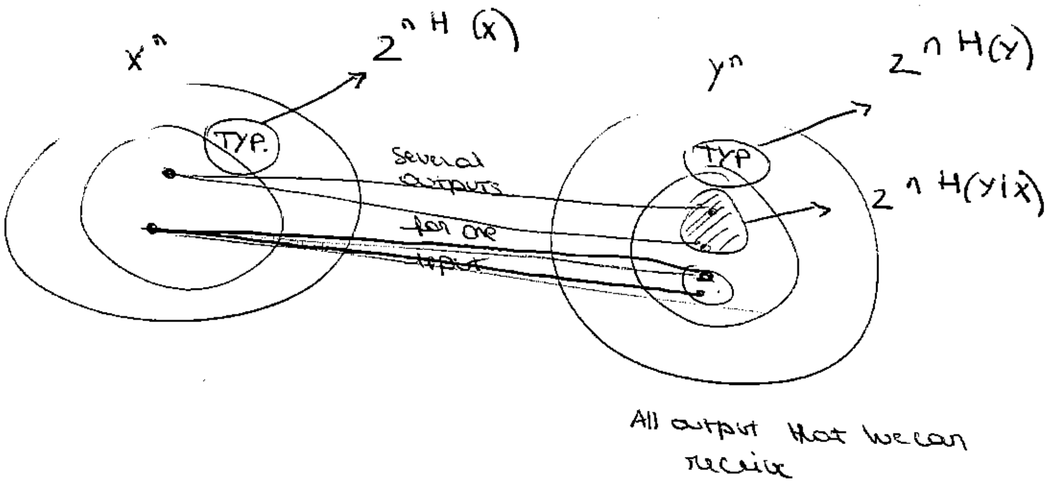
SHANNON 2° THEOREM (NOISY CHANNEL)

$$C_{OPER} \equiv C_{INFO}$$



$$P_e = \mathbb{P}[\hat{W} \neq W] \leq \varepsilon$$





$$\text{Nr of nonoverlapping ensembles} \leq \frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{n(H(Y) - H(Y|X))} = 2^{nI(X:Y)}$$

$$\text{Nr of distinguishable messages} \leq 2^{nI(X:Y)}$$

Info sent per use of channel $= \frac{1}{n} \log_2 [\text{Nr of msg}] \leq \underline{I(X:Y)}$

$\uparrow$

$\max_{p(x)} \Rightarrow C$

<u>PROOF</u> RANDOM CODE & DECODING BASED ON JOINT TYPICALITY.

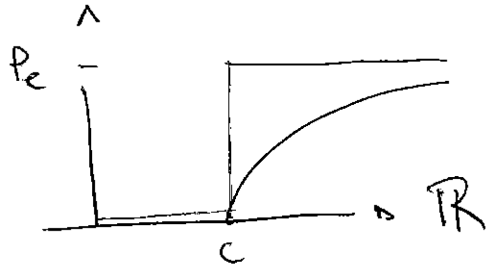
<u>Theorem</u> If $R \leq C$, they $\exists$ code such that $\mathbb{P}_e$ arbitrary small $\leq \varepsilon$ for asymptotically large n

$$\lim_{n \to \infty} \mathbb{P}_e^{(n)} = 0$$

<u>WEAK CONVERSE</u> If $\lim_{n \to \infty} \mathbb{P}_e^{(n)}$, then $R \leq C$

If $R > C$, then $\lim_{n \to \infty} \mathbb{P}_e^{(n)} = P > 0$

<u>STRONG CONVERSE</u> if $R > C$, then $\lim_{n \to \infty} \mathbb{P}_e^{(n)} = 1$



Collège Saint-Julien
Ath

Nom ....

Prénom ....

Classe .... Date ....

Branche .... Professeur ....

Chapter 5 : Error correcting codes

Binary Symmetric Channel

$$p(y|x)$$

$$\boxed{ENC} \xrightarrow{w} \boxed{BSC} \longrightarrow v \;\boxed{DEC}$$

$$\underbrace{0110\ldots1}_{n} \qquad 0 < p < \tfrac{1}{2} \qquad \underbrace{011\ldots1}_{n \text{ bits}}$$

M codewords of length n

$M \leq 2^n$

received sequence

(2^n possible seq.)

"Ideal decoder" : associates to received sequence v the codeword w that is the most probable.

Assumption : M code words are equiprobable $p(w) = \frac{1}{M}$

For fixed v, choose w by $\max_w p(w|v)$

Bayes rule $p(w|v) \, p(v) = p(v|w) \, p(w)$

$$\underset{w}{\max}\; p(w|v) = \underset{w}{\max}\; p(v|w) \cdot \underbrace{\frac{1}{M} \cdot \frac{1}{p(v)}}_{\text{fixed}}$$

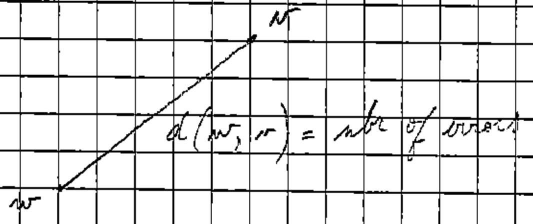
"Ideal decoder" : associates to v the codewords that maximizes $p(v|w)$

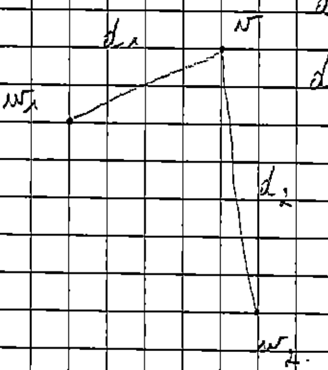
Hamming distance $d(v_1, v_2)$ = nbr of bits that are different in v_1 and v_2

1) $d(v_1, v_2) \geq 0$ $\quad d(v_1, v_2) = 0$ iff $v_1 = v_2$

2) $d(w_1, w_2) = d(w_2, w_1)$

3) $d(w_1, w_2) \leq d(w_1, w_3) + d(w_3, w_2)$





$d_1 = d(w_1, v)$

$d_2 = d(w_2, v)$

$$\max_w p(v|w) = p^{d(w,v)} \cdot (1-p)^{n-d(w,v)}$$

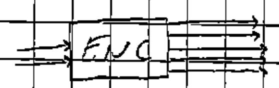
$$\frac{p(v|w_1)}{p(v|w_2)} = \frac{p^{d_1}(1-p)^{n-d_1}}{p^{d_2}(1-p)^{n-d_2}} = \left(\frac{p}{1-p}\right)^{d_1-d_2} = \underbrace{\left(\frac{1-p}{p}\right)}_{\geq 1}{}^{d_2-d_1}$$

$$p(v|w_1) > p(v|w_2) \quad \text{iff} \quad d_1 < d_2$$

choose w_1

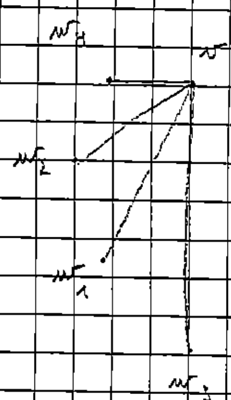
"Ideal decoder" $\equiv$ "Minimum-distance decoder"

<u>Example</u> $M = 4$



$$\begin{cases} w_1 = 00000 & d_1 = 3 \\ w_2 = 10011 & d_2 = 2 \\ w_3 = 11100 & d_3 = 4 \\ w_4 = 01111 & d_4 = 1 \end{cases}$$

$v = 01011$



Collège Saint-Julien
Ath

Nom

Prénom

Classe Date

Branche Professeur

Relation between distance of a code and error correction

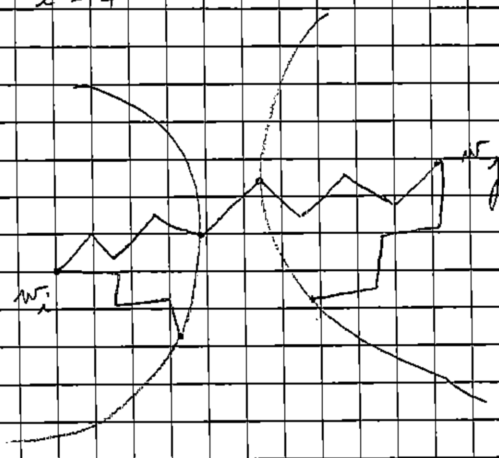
Theorem 1: Given a set of code words $w_1, w_2, \dots, w_n$ of length n satisfying $d(w_i, w_j) \geq \underbrace{2e+1}_{\text{odd integer}}$ $\forall i \neq j$ e = integer > 0

& converse then code corrects up to e errors.

Theorem 2: Given a set of code words $w_1, w_2, \dots, w_n$ of length n satisfying $d(w_i, w_j) \geq \underbrace{2e}_{\text{even integer}}$ $\forall i \neq j$ e = integer > 0

& converse then code corrects up to $e-1$ errors and detects e errors.

Proof: $e = 4$



$d = 2e + 1 = 9$

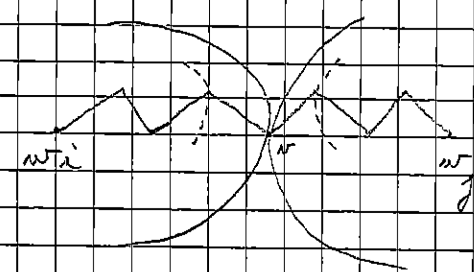
Sphere $= \{x : d(w_i, x) \leq 4\}$

. Sufficient to have $d = 2e+1 = 9$ in order to correct $e = 4$ errors

it is necessary to have 5 errors to have the risk of decoding error.

Necessary to have $d = 2e+1 = 9$ in order to correct $e=4$ errors.

If distance $d = 2e = 8$



$d = 2e = 8$

$$\underbrace{d(w_i, w_j)}_{2e} \leq \underbrace{d(w_i, v)}_{e} + d(v, w_j)$$

$\Rightarrow d(v, w_j) \geq e = 4 \Rightarrow$ cannot correct $e=4$ errors

(can still correct $e-1=3$ errors)

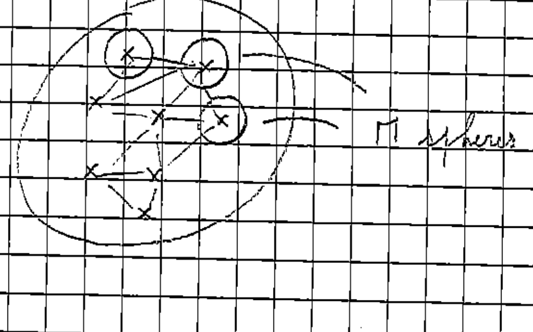
Nevertheless, I can detect $e=4$ errors

Parameters characterizing a ECC

$$\begin{cases} M = \text{nbr of codewords} \\ n = \text{length of codewords} \\ d = \min\limits_{i \neq j} d(w_i, w_j) \end{cases}$$

$\longrightarrow$ (minimum) distance of the code

Hamming's bound



* $R = \dfrac{\log M}{n}$ bits / use of channel

$\Rightarrow M = 2^{nR}$

$\Rightarrow$ many codewords

* $d \geq 2e+1$

$\Rightarrow$ codewords distant from each other

Collège Saint-Julien
Ath

Nom ....

Prénom ....

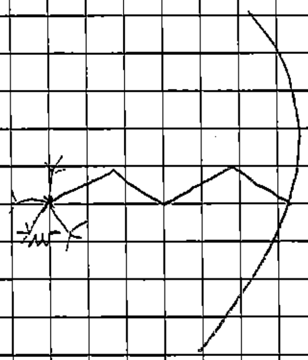
Classe .... Date ....

Branche .... Professeur ....

How many codewords can we have if e errors can be corrected ?

Hamming's bound $\boxed{M \leq \dfrac{2^n}{\sum_{i=0}^{e} \binom{n}{i}}}$

$$\binom{n}{i} = \frac{n!}{(n-i)!\,i!}$$



Nb of sequences inside sphere

$$= 1 + n + \binom{n}{2} + \dots$$

$$= \sum_{i=0}^{e} \binom{n}{i}$$

Sphere must be non-overlapping

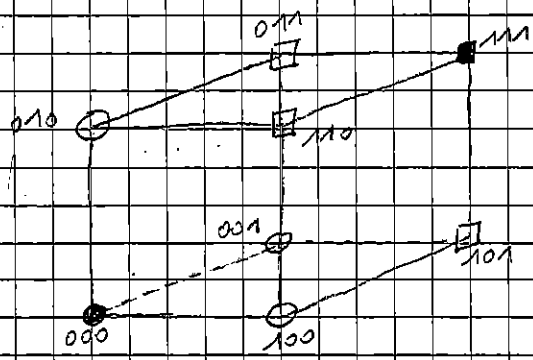
$$M \cdot \sum_{i=0}^{e} \binom{n}{i} \leq 2^n \qquad \text{necessary condition}$$

Ex. repetition code $0 \rightarrow 000 \Rightarrow \begin{cases} 001 \\ 010 \\ 100 \end{cases} \Rightarrow$ majority rule $\Rightarrow 0$

$M = 2$ $\quad 1 \rightarrow 111 \Rightarrow \begin{cases} 110 \\ 101 \\ 011 \end{cases} \Rightarrow$ majority rule $\Rightarrow 1$

$n = 3$

$d = 3 \Leftrightarrow$ correct $e = 1$ error $\qquad$ Hamming $M = 2 \overset{?}{\leq} \dfrac{2^n = 2^3 = 8}{1+n = 4} = 2$



Hamming's bound: necessary ~~sufficient~~ condition for code to correct e errors

$$M \le \frac{2^n}{\sum_{i=0}^{e}\binom{n}{i}} \Rightarrow \boxed{M \le \left\lfloor \frac{2^n}{\sum_{i=0}^{e}\binom{n}{i}} \right\rfloor}$$

Example: $\begin{cases} n = 4 \\ e = 1 \end{cases}$ $\hookrightarrow d = 3$

$$M \le \left\lfloor \frac{2^4}{1+4} \right\rfloor = \left\lfloor \frac{16}{5} \right\rfloor = 3 \qquad \underline{\underline{M \le 2}}$$

$$\begin{cases} w_1 = 0\,0\,0\,0 \\ w_2 = 0\,1\,1\,1 \\ w_3 = 1\,\boxed{?\,?\,?} \end{cases} \quad d_{13} = 1 + x \quad d_{23} = 1 + (3 - x) = 4 - x \quad d_{12} = 3$$

x = nbr of "1"

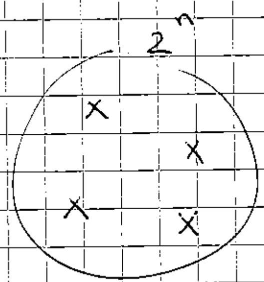
	d_{12}	d_{13}	d_{23}
$x=0$	3	1	4
1	3	2	3
2	3	3	2
3	3	4	1

Coding theory

- ERROR CORRECTING CODES

ENC $\xrightarrow{w}$ BSC $\xrightarrow{v}$

w codeword, v received sequence



$$\begin{cases} M = \text{no of codewords} \\ n = \text{length of codewords} \end{cases}$$

d = distance
$\hookrightarrow \min_{i \neq j} d(w_i, w_j)$

$d = 2l + 1$

- Hamming's bound $\quad M \leq \left\lfloor \dfrac{2^n}{\sum_{i=0}^{l} \binom{n}{i}} \right\rfloor$

Ex: $n = 5$

① $l = 1 \quad M \leq \left\lfloor \dfrac{32}{1+5} \right\rfloor = 5 \dashrightarrow M \leq 4$

$\xrightarrow{2}$ ENC $\xrightarrow{5}$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{matrix} \}3 \\ \}4 \\ \}3 \end{matrix}$$

3

$R = \dfrac{2}{5}$

② $l = 2 \quad M \leq \left\lfloor \dfrac{32}{1+5+\binom{5}{2}} \right\rfloor = 2$

$\left.\begin{matrix} 0\,0\,0\,0\,0 \\ 1\,1\,1\,1\,1 \end{matrix}\right) 5$

$\binom{5}{2} = \dfrac{5 \cdot 4}{2 \cdot 1} = 10$

$\xrightarrow{1}$ ENC $\xrightarrow{5}$

Repetition code

$R = \dfrac{1}{5}$

- HAMMING'S CODES

$\xrightarrow{n}$ DEC $\rightarrow$ most probable codeword

2^n entries

- Parity bit $\underbrace{0\ 1\ 0\ 1\ 1}_{5\text{ bit}}\ \underbrace{1}_{\text{parity bit}}$

Ex: $\underbrace{\begin{matrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{matrix}}_{\text{information bits}}\ \underbrace{\begin{matrix} 0 \\ 1 \\ 1 \\ 0 \end{matrix}}_{\text{parity check bit}} \left(\begin{matrix} \to 2\text{"0"} \\ \to 1\text{"1"} \end{matrix}\right)$ + Data $\left\{\begin{matrix} M = 4 \\ n = 3 \\ d = 2 \end{matrix}\right.$ → detect 1 → error, correct 0 error

- $w = x_1 x_2 \dots x_n \Rightarrow x_1 + x_2 + \dots + x_n = 0 \bmod 2$

$(x_i = 0, 1)$

- $w = (x_1 \dots x_n)^T$ is CODEWORD IFF

$$m \text{ equations} \left\{\begin{matrix} h_{11} x_1 + h_{12} x_2 + \dots + h_{1n} x_n = 0 \\ \vdots \\ h_{m1} x_1 + h_{m2} x_2 + \dots + h_{mn} x_n = 0 \end{matrix}\right. \qquad h_{ij} = 0, 1 \quad H = [h_{ij}]$$

$$\Longleftrightarrow m \text{ rows} \left\{ \underbrace{\begin{pmatrix} h_{11} & \dots & h_{1n} \\ \vdots & & \vdots \\ h_{m1} & \dots & h_{mn} \end{pmatrix}}_{n \text{ columns}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 0 \right. \iff \underline{H_w = 0}$$

Hamming's matrix
priority check matrix

$t = \text{rank } H \leq m < n$

{ m linearly indep rows
m linearly indep columns

$$m \left\{ \begin{pmatrix} h_{1,1} & \dots & h_{1,m} & \vdots & h_{1,m+1} & \dots & h_{1,n} \\ \vdots & & & \vdots & & & \vdots \\ h_{m,1} & \dots & h_{m,m} & \vdots & h_{m,m+1} & \dots & h_{m,n} \end{pmatrix} \right. \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \hline x_{m+1} \\ \vdots \\ x_n \end{pmatrix} = 0$$

m linearly columns

m parity check bits | n − m info bits

$$m \text{ eq} \left\{ \begin{array}{l} h_{11} x_1 + \dots + h_{1,m} x_m = + h_{1,m+1} + \dots + h_{1,n} x_n \\ \vdots \\ \underbrace{h_{m,1} x_1 + \dots + h_{m,n} x_m}_{} = + h_{m,m+1} \underbrace{x_{m+1} + \dots + h_{m,n} x_n}_{n-m \text{ INFO BITS}} \end{array} \right.$$

Solve for n parity bits.

$$w = (\underbrace{x_1 \; x_2 \; \dots \; x_n}_{m \text{ parity BITS}} \mid \underbrace{x_{m+1} \; \dots \; x_n}_{n-m \text{ info BITS}}) \rightarrow k = n-m$$
$$\rightarrow M = 2^k$$

$$R = \frac{\log_q M}{n} \longleftrightarrow M = 2^{nR}$$

$\xrightarrow{k}$ ENC $\xrightarrow{n}$ n physical: [m (parity) | k (info)] $\Rightarrow R = \frac{k}{n} = \frac{\text{nb of log bits}}{\text{nb of phys bits}}$

Example of Hamming's CODE

$$\left\{ \begin{array}{l} n = 6 \\ m = 3 \quad (= \text{rank } H) \\ k = n - m = 3 \quad \text{info bits} \end{array} \right.$$

$$H \begin{pmatrix} x_1 \\ \vdots \\ x_3 \\ \hline x_4 \\ \vdots \\ x_6 \end{pmatrix} = 0$$

$$H = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{array} \right)$$

3 bit parity   3 info bits

résolution du système

Fix:
$$\left\{ \begin{array}{l} x_4 = 0 \\ x_5 = 1 \\ x_6 = 1 \end{array} \right. \Rightarrow \begin{array}{l} x_1 + 0 + 1 + 0 = 0 \\ x_2 + x_3 + 0 + 0 + 1 = 0 \\ x_1 + x_3 + 0 + 0 + 1 = 0 \end{array} \left\{ \begin{array}{ll} x_1 = 1 & \\ x_2 + x_3 = 1 & x_2 = 1 \\ x_1 + x_3 = 1 & x_3 = 0 \end{array} \right.$$

$$\Rightarrow w = (\underbrace{1\,1\,0}_{\text{parity}} \mid \underbrace{0\,1\,1}_{\text{info}})$$

List of codewords $\quad n = 2^k = 8$

$w_0 = 0\,0\,0 \quad 0\,0\,0$ $\quad d=2$

$w_1 = 0\,0\,1 \quad 0\,0\,1$ $\quad d=2$

$w_2 = 1\,1\,1 \quad 0\,1\,0$

$w_3 = 1\,1\,0 \quad 0\,1\,1$

$\vdots$

$w_7 = 0\,0\,0 \quad 1\,1\,1 \qquad d=2$ detects 1 error.

- Hamming's code is linear

$$\left.\begin{array}{l} w_1 = 001\ 001 \\ w_2 = 111\ 010 \end{array}\right\} w_1 + w_2 = 110\ 011 = w_3$$

w_i, w_j codewords $\Leftrightarrow \left\{\begin{array}{l} H w_i = 0 \\ H w_j = 0 \end{array}\right. \Rightarrow (w = w_i + w_j)$ is a codeword

$$\Rightarrow H w = H(w_i + w_j) = H w_i + H w_j$$

- (min) distance $d = \min\limits_{i \neq j} d(w_i, w_j)$ $\qquad \boxed{d = W}$

(min) weight $W = \min\limits_{k \neq 0} W(w_k)$

$$d = \min_{i \neq j} d(w_i, w_j) = \min_{\substack{i \neq j \\ k \neq 0}} W(w_i + w_j) = \min_{k \neq 0} W(w_k) = W$$

- Decoding $\quad w \rightarrow \boxed{BSC} \xrightarrow{v} \boxed{DEC}$

$w = (x_1 \dots x_n)^T ; \; v = (y_1 \dots y_n)^T \qquad C \equiv H v$

correcting vector

syndrome [illegible]

- NO ERROR $(v=w) \Rightarrow C=0$
- WAS ERROR(S) $\Leftarrow C \neq 0$

$\not\Rightarrow$

- $V = W + Z$ $\quad z = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \begin{array}{l} \\ \\ \\ \leftarrow j_1 \\ \\ \leftarrow j_2 \\ \\ \\ \leftarrow j_e \end{array}$

error pattern vector

- $C = Hv = H(w+z) = \cancel{Hw} + Hz = Hz$

$= \sum_{j \in \{j_1, j_2, \dots, j_e\}} \text{column}_j(H)$

EXAMPLE:

$$H = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{array}\right)$$

3 parity bits $\quad$ 3 info bits

w codewords $\Leftrightarrow Hw = 0$

- Error of bit #1 $\quad C = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Error of bits #2, #3, #5 $\quad C = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Single error of bit #j $\Rightarrow C = \text{column}_j(H)$

- Hamming's code corrects 1 error iff all columns are distinct

2 error iff all ensembles of 4 colums are lin indep

3 errors iff " " 6 columns

" "

(4 colonnes par le set de toutes les colonnes)

- "CANONICAL" HAMMING'S CODE

$$H = \begin{array}{c} \begin{array}{cccc|ccc} \text{INFO} & & & & \text{PARIBITY} & & \end{array} \\ \left(\begin{array}{cccc|ccc} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right) \\ \begin{array}{ccccccc} \uparrow & & & & & & \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \end{array}$$

$0\ 0\ 1 \leftarrow 1$

$n = 7$

$m = 3$ parity bits $\quad k = n - m = 4$ info bits

$w_0 = 0\,0\,0\,0\ \ 0\,0\,0$

$w_1 = 0\,0\,0\,1\ \ 1\,1\,1$ $\quad d = 3 = W \Rightarrow e = 1$

$w_2 = 0\,0\,1\,0\ \ 1\,1\,0$

$\vdots$

$w_{15} = 1\,1\,1\,1\ \ 1\,1\,1$

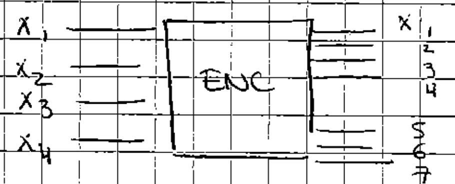
$$\rightarrow \begin{cases} x_5 + x_6 + x_7 = x_4 \\ x_6 + x_7 = x_2 + x_3 \\ x_5 + x_7 = x_1 + x_3 \end{cases}$$

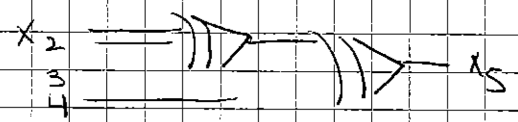
$M = 2^4 = 16$

$x_6 = x_1 + x_3 + x_4$

$x_7 = x_1 + x_2 + x_4$

$x_5 = x_2 + x_3 + x_4$





Decoding

$y \rightarrow \boxed{H} \rightarrow C = Hy = Hz = H \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \text{col}_j(H) = j$ error in bit #j

$$\begin{cases} C_1 = y_4 + y_5 + y_6 + y_7 \\ C_2 = y_2 + y_3 + y_6 + y_7 \\ C_3 = y_1 + y_3 + y_5 + y_7 \end{cases}$$

Coding Theory

19/5/14.

- Hamming's code

$$\underbrace{\begin{pmatrix} h_{11} & h_{1n} \\ h_{m1} & h_{mn} \end{pmatrix}}_{H} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 0$$

$$\operatorname{rank} H = m < n$$
$$k = n - m$$
$$M = 2^k$$

$$\begin{aligned} C &= Hv \\ &= H(w+z) \\ &= Hz \end{aligned}$$

$Hw = 0 \iff w$ codeword

- Canonical code

$$H = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}}_{7} \Big\} 3$$

- $M \le \dfrac{2^n}{\sum_{i=0}^{e} \binom{n}{i}}$

$$2^4 \overset{?}{\le} \frac{2^7}{1+7 = 2^3} = 2^4$$

$k = 4$
$M = 16$
$e = 1$

$$\begin{cases} m \\ n = 2^m - 1 \\ k = 2^m - 1 - m \\ e = 1 \end{cases}$$

- $2^{n-m} = 2^k = \dfrac{2^n}{1+n} \longrightarrow 1 + n = 2^m$

$$\Rightarrow n = 2^m - 1 \quad \begin{cases} m = 3 \\ n = 2^3 - 1 = 7 \end{cases}$$

- $m = 4$, $n = 15$ $\quad \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \Rightarrow e = 1$

- Ideal decoder

- $\longrightarrow \boxed{\text{BSC}} \xrightarrow{n} v \Rightarrow$ Table with 2^n entries $\longrightarrow$ Most probable w

- $\longrightarrow \boxed{\text{BSC}} \xrightarrow{n} v \Rightarrow \boxed{H} \xrightarrow{m} C \longrightarrow$ Table with 2^m entries $\longrightarrow$ Most probable z

- $m \ll n$
  $2^m <<< 2^n$

Lower BOUNDS on Nr of Parity check BITS.

n bits

k info   m parity

$k\uparrow \quad M = 2^k \uparrow \quad R\uparrow$

$m\uparrow \quad e\uparrow$

① Hamming's bound

n, e fixed

$$\boxed{2^m \geq \sum_{i=0}^{e} \binom{n}{i}}$$

- Vector C may take 2^m values.
- Nr of error patterns $= 1 + n + \binom{n}{2} + \dots + \binom{n}{e}$

- "General" Hamming's bound

  - $$M \leq \frac{2^n}{\sum_{i=0}^{e} \binom{n}{i}}$$

  - $$2^k = 2^{n-m} \leq \frac{2^n}{\Sigma \dots}$$

$$\longrightarrow \sum_{i=0}^{e} \binom{n}{i} \leq 2^m$$

→ Necessary ~~Sufficient~~

$$\begin{cases} n = 10 \\ e = 2 \end{cases} \qquad 2^m \geq 1 + 10 + \overset{45}{\binom{10}{2}} - 56$$

$$\underline{\underline{m \geq 6}}$$

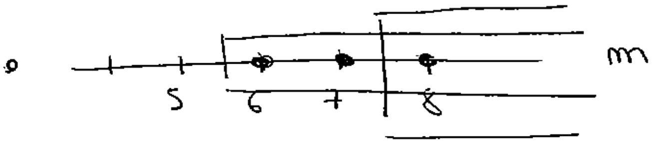
② Gilbert Varshamov bound

- n, e fixed $\quad 2^m > \sum_{i=0}^{2e-1} \binom{n-1}{i}$

$$H = \overbrace{\left(\ ////\ \right)}^{n} \Big\} m$$

- Sufficient ~~Necessary~~

$$\begin{cases} n = 10 \\ e = 2 \end{cases} \qquad 2^m > 1 + \overset{9}{\binom{9}{1}} + \overset{36}{\binom{9}{2}} + \overset{84}{\binom{9}{3}} = 130 \qquad m \geq 8$$



$\nwarrow m = 7 \; , \; k = 3$.

- $m_{GV} \geq m_H$

$$\frac{m_{GV}}{m_H} \leq 2$$

$\underline{\underline{e=1}}$ H: $2^m \geq 1 + n$

GV: $2^m > 1 + (n-1) = n$

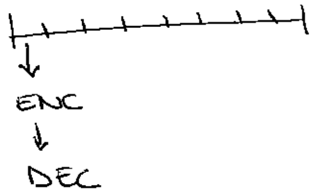
$\Big\} \; 2^m \geq 1 + n$

$\hookrightarrow$ necessary & sufficient condition to correct $e=1$ error.

OVERVIEW OF ECC.

- "Block" CODES (m, n)
  - Linear codes
    - Hamming's Code
      - cyclic codes
        - BCH codes
        - Reed Solomon codes.
      - Non cyclic codes
  - Non linear codes

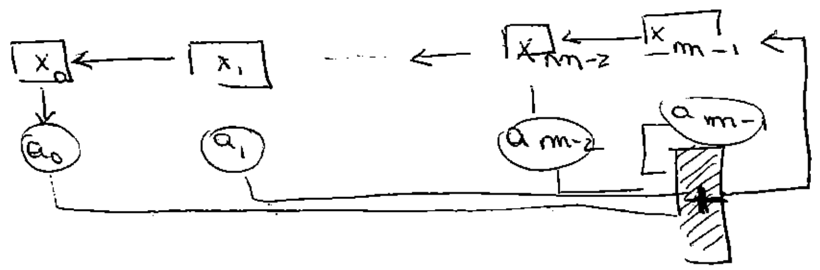
Trellis Code



- Linear codes
- non Linear

CYCLIC CODES:

SHIFT Register
+ Feedback



$$
\begin{cases}
x_0' = x_1 \\
x_1' = x_2 \\
\vdots \\
x_{m-2}' = x_{m-1} \\
x_{m-1}' = a_0 x_0 + a_1 x_1 + \dots + a_{m-1} x_{m-1}
\end{cases}
\qquad \boxed{X' = T X}
$$

transition Matrix

$$
X = \begin{pmatrix} x_0 \\ \vdots \\ x_{m-1} \end{pmatrix} ; \quad X' = \begin{pmatrix} x_0' \\ \vdots \\ x_{m-1}' \end{pmatrix} ; \quad T = \begin{pmatrix} 0 & 1 & 0 & & \\ & 0 & 1 & 0 & \\ & & 0 & \ddots & \\ & & & 0 & 1 \\ a_0 & a_1 & \dots & & a_{m-1} \end{pmatrix}
$$

$\det T = a_0 = 1$.

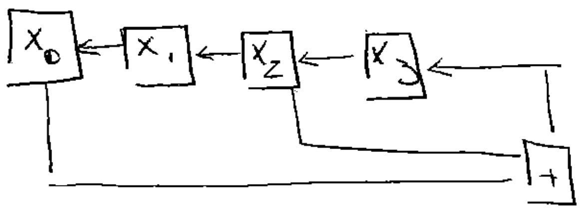
- Time Evaluation

$X_0 \longrightarrow X_1 = T X_1 = T^2 X_0 \curvearrowright \dots \curvearrowright X_0$

- $X_0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \circlearrowleft 1$

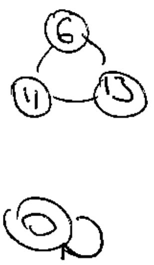
- Cycle whose period is $\leq 2^m - 1$

Ex :
$$
\begin{cases}
a_0 = 1 \\
a_1 = 0 \\
a_2 = 1 \\
a_3 = 0
\end{cases}
$$



3

4

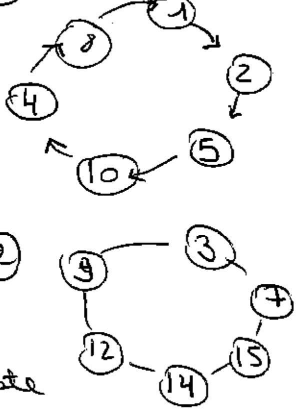


0	0	0	1
0	0	1	0
0	1	0	1
1	0	1	0
0	1	0	0
1	0	0	0
0	0	0	1

initial state

1

2



DEF : code is CYCLIC when it FULFILLS Property

IF : $w = (w_0\, w_1\, \dots\, w_{n-1})^T$ is codeword

THEN : $w' = (w_{n-1}\; w_0\, w_1 \dots w_{n-2})^T$ is codeword!

Theorem : IF one constructs Hamming's code as follows

$$H = \underbrace{\left(X \mid TX \mid T^2X \mid \dots T^{n-1}X \right)}_{n} \Big\}\, n$$

• columns = subsequent states over one full cycle

n = period of cycle

then resulting code is cyclic

X = some initial state $= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$n = 6$

$m = 4$

$k = 2 \Rightarrow M = 4$

codewords : $\begin{cases} w_0 = 000\ 000 \\ w_1 = 101\ 010 \\ w_2 = 010\ 101 \\ w_3 = 111\ 111 \end{cases}$

Proof : IF $w = (w_0 \dots w_{n-1})^T$ codeword

$\to Hw = 0$

$\to (X \mid TX \mid \dots \mid T^{n-1}X) \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{n-1} \end{pmatrix} = 0$

$\to w_0 X + w_1 TX + \dots w_{n-2} T^{n-2} X + w_{n-1} T^{n-1} X = 0$

$w_0 TX + w_1 T^2X + \dots + w_{n-2} T^{n-1} X + w_{n-1} X = 0$ T

$\to (X \mid TX \mid \dots T^{n-1}X) \begin{pmatrix} w_{n-1} \\ w_0 \\ w_1 \\ \vdots \\ w_{n-2} \end{pmatrix} = 0 \to w' = (w_{n-1}\, w_0\, w_1 \dots w_{n-2})^T$ is a codeword

CLASS OF CYCLIC CODES

$$H = \left(X \mid TX \mid \cdots \mid T^{n-1}X \right) \text{ over a cycle}$$

$$T = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ a_0 & a_1 & \cdots & a_{n-1} \end{pmatrix} ; \quad X = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Pour INFO

- Ex : BOSE - CHAUDHURI - HOCQUENGHEM CODES.

e, $T(q \times q)$, $n = 2^q - 1$

$$H = \left(\begin{array}{c|c|c|c|c} X & TX & T^2X & \cdots & T^{n-1}X \\ \hline X & T^3X & T^5X & & T^{3(n-1)}X \\ \hline X & T^5X & T^{10}X & & T^{5(n-1)}X \\ \hline \vdots & & & & \\ X & T^{(2e-1)}X & T^{4e-2}X & \cdots & T^{(2e-1)(n-1)}X \end{array} \right) \begin{array}{l} \left.\vphantom{\begin{array}{c}X\\X\\X\\X\end{array}}\right\} e \text{ blocs of } q \text{ rows.} \\ \left.\vphantom{X}\right\} q \end{array}$$