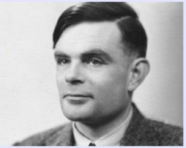
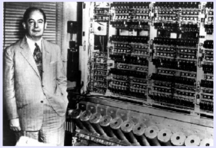
  

Information, Calcul et Communication

Compléments de cours

J.-C. Chappelier

EPFL

Combient vaut

1000×342000?

$\sqrt{2}$

$1/3$

$3x = 1$

Leçon II.3 – Examen final 2018 Q1.1

$$-\sum p_i \log p_i = -\sum \frac{n_i}{N} \log \frac{n_i}{N} \qquad p_i = \frac{n_i}{N}$$

Un sac contient 24 billes de 4 couleurs différentes :
12 billes rouges, 8 billes bleues, 2 billes vertes, et 2 billes oranges.

Quelle est, en bit, l'entropie du jeu consistant à deviner la couleur d'une bille tirée au hasard ?

Donnez votre réponse sous la forme $a + b\log_2(3)$.

$$\underbrace{\log 24}_{3+\log 3} - \frac{1}{24}\left(12 \underbrace{\log 12}_{2+\log 3} + 8 \underbrace{\log 8}_{3} + 4 \underbrace{\log 2}_{1}\right)$$

$\hookrightarrow 1 - \frac{1}{8} + \frac{1}{2}\log 3$

a $\quad\quad$ b

$5/6$ $\quad\quad$ $1/2$

$0\,1\,0\,1\,0\,1$

$X = 0\;1\;0\;0\;1\;\ldots$ $\quad\quad$ $X_i \in \{0,1\}$

Leçon II.3 – Examen final 2018 Q1.1

Un sac contient 24 billes de 4 couleurs différentes :
12 billes rouges, 8 billes bleues, 2 billes vertes, et 2 billes oranges.

Quelle est, en bit, l'entropie du jeu consistant à deviner la couleur d'une bille tirée au hasard ?

Donnez votre réponse sous la forme $a + b\log_2(3)$.

$$\frac{1}{2} + \frac{1}{3}\log_2(3) + \frac{1}{6}\log_2(12) = \mathbf{\frac{5}{6}} + \mathbf{\frac{1}{2}}\log_2(3)$$

Leçon II.3 – Examen 2 2016 Q5

N

X :

0 1, 0 1, 0 1,

On utilise une représentation des nombres entiers *signés* sur N bits et on s'en sert pour représenter uniquement les nombres entiers positifs (zéro y compris). Sachant que N est pair, quelle est (en bit) l'entropie maximale (telle que définie en cours) pour une telle séquence de symboles 0/1 ?

A] $\log_2(N-1)$ **B]** $\log_2(N)-1$ **C]** 1 **D]** $N-1$

$\hookrightarrow h(1)$

$0 \leq H(X) \leq 1$ bit

et si N est impair ?

$\overbrace{0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0}^{N}$

$$\left.\begin{array}{ll} \text{nb de } 0 : & \dfrac{N+1}{2} \\[2ex] \text{nb de } 1 : & \dfrac{N-1}{2} \end{array}\right\} \text{ binaire} : h(p_0)$$

$$h\left(\frac{1}{2}+\frac{1}{2N}\right)$$

$$p_0 : \frac{N+1}{2N} = \frac{1}{2} + \frac{1}{2N}$$

Leçon II.3 – Examen 2 2016 Q5

On utilise une représentation des nombres entiers *signés* sur N bits
et on s'en sert pour représenter uniquement les nombres entiers positifs (zéro y compris).
Sachant que N est pair, quelle est (en bit) l'entropie maximale (telle que définie en cours)
pour une telle séquence de symboles 0/1 ?

A] $\log_2(N-1)$ **B]** $\log_2(N)-1$ ***C]** 1 **D]** $N-1$

Leçon II.3 – Examen final 2018 Q3.1

F G

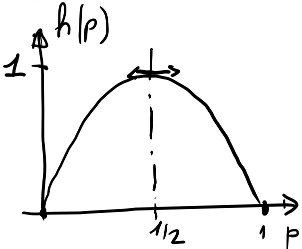
p 1−p

Quelle est l'entropie (telle que définie en cours) d'un mot X constitué de n fois la lettre 'F' et m fois la lettre 'G' : « $\underbrace{\texttt{F......F}}_{\longleftarrow n \longrightarrow}\underbrace{\texttt{G......G}}_{\longleftarrow m \longrightarrow}$ » ?

Soit $p = \dfrac{n}{n+m}$.

Utilisez p pour exprimer cette entropie comme une fonction de p uniquement :

$$H(X) = -p\log p - (1-p)\log(1-p) = h(p)$$



Leçon II.3 – Examen final 2018 Q3.1

Quelle est l'entropie (telle que définie en cours) d'un mot X constitué de n fois la lettre 'F' et m fois la lettre 'G' : « $\underbrace{\texttt{F......F}}_{n}\underbrace{\texttt{G......G}}_{m}$ » ?

Soit $p = \dfrac{n}{n+m}$.

Utilisez p pour exprimer cette entropie comme une fonction de p uniquement :

$$H(X) = -\frac{n}{n+m}\log_2\left(\frac{n}{n+m}\right) - \frac{m}{n+m}\log_2\left(\frac{m}{n+m}\right) = \underbrace{p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right)}_{h(p)}$$

Leçon II.3 – Examen 2 2016 Q6

Considérons une séquence de caractères de la forme « $\underbrace{AB******}_{\leftarrow 6 \rightarrow}$ », $= X$

$n=4$

où le symbole '$*$' peut être remplacé par n'importe quelle lettre de l'alphabet $\{A, B, C, D\}$.

Quelle bornes (en bit) *les plus strictes* pouvez-vous donner pour l'entropie H d'une telle séquence ?

$$-\frac{1}{6}\log 6 - \frac{5}{6}\log\frac{6}{5}$$

$$-\frac{7}{8} \cdots\cdots \frac{1}{8}$$

$$h\left(\frac{1}{4}\right)$$

$$1$$

$$h\left(\frac{1}{8}\right) \leq H(X) \leq 2$$

$$\log_2 4$$

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

X_{MAX}

ABCDABCD

~~ABABABAB~~

$H(X_{MAX}) = 2$

X_{min}

ABAAAAAA $\quad P_B = \frac{1}{8}$

Leçon II.3 – Examen 2 2016 Q6

Considérons une séquence de caractères de la forme « $AB\underset{\leftarrow 6 \rightarrow}{******}$ »,

où le symbole '$*$' peut être remplacé par n'importe quelle lettre de l'alphabet $\{A, B, C, D\}$.

Quelle bornes (en bit) *les plus strictes* pouvez-vous donner pour l'entropie H d'une telle séquence ?

$$h(\frac{1}{8}) \leq H(X) \leq 2 \text{ bit}$$

avec :

$$\begin{aligned} h(\frac{1}{8}) &= \frac{1}{8}\log_2 8 + \frac{7}{8}\log_2\frac{8}{7} \\ &= \frac{3}{8} + \frac{21}{8} - \frac{7}{8}\log_2 7 \\ &\simeq 0.54 \text{ bit} \end{aligned}$$

Leçon II.3 – Examen 2 2016 Q4

Soit H_1 l'entropie d'une séquence (non vide) de lettres S_1 et H_2 l'entropie d'une séquence de lettres S_2 contenant strictement S_1
(c.-à-d. que la séquence S_1 en tant que telle est une sous-séquence de S_2).

A] On a forcément $H_2 > H_1$.

B] On a forcément $H_2 < H_1$.

C] On a forcément $H_2 = H_1$.

D] Aucune des trois autres propositions.

Leçon II.3 – Examen 2 2016 Q4

Soit H_1 l'entropie d'une séquence (non vide) de lettres S_1 et H_2 l'entropie d'une séquence de lettres S_2 contenant strictement S_1
(c.-à-d. que la séquence S_1 en tant que telle est une sous-séquence de S_2).

A] On a forcément $H_2 > H_1$.

B] On a forcément $H_2 < H_1$.

C] On a forcément $H_2 = H_1$.

*__D]__ Aucune des trois autres propositions.

Leçon II.4 (compression) – Points clés

- algorithme (de compression) de Shannon-Fano
- théorème de Shannon (de compression sans perte)
- lien avec la définition intuitive de l'entropie
  (nombre moyen de questions)
- algorithme (de compression) de Huffman
- optimalité (compression sans perte) des codes de Huffman

th. Shannon

$\forall$ X une source :

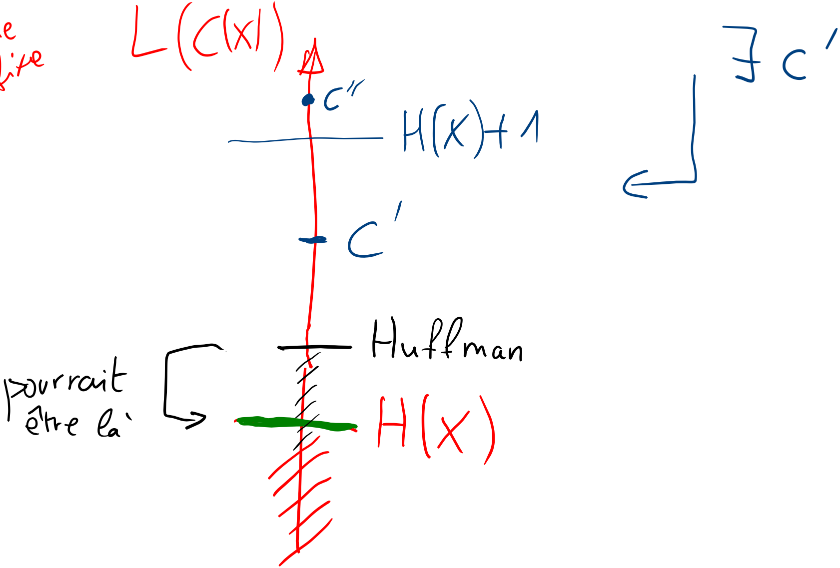
① $\forall$ $C(X)$ code sans-préfixe et sans-perte

$$H(X) \leq L(C(X))$$ long. moyenne

② $\exists$ $C'(X)$ $$L(C'(X)) < H(X) + 1$$

code .....

$\forall C$ sans-perte
sans-préfixe



Leçon II.4 (compression) – Étude de cas

01100011 01111010

→ 101 → 11

Vous souhaitez envoyer un message, qui,

- ~~considéré bit à bit, a une entropie de 0.9 bit,~~
- et considéré octet par octet, une entropie de 5.51 bit.

Sachant que ce message (d'origine, complet) a une taille de 10 Ko, quelle taille pouvez-vous espérer après compression (sans autre information que celles fournies ici) ?

0 : ~~AA~~ { 0 1

1 : ~~BCD~~ { 1 0

coder : symbole → code(symbole)

octet → code

$\forall$ code C des octets :

$$5.51 \leq L_C$$

$\exists$ (p. ex Huffman) code C' des octets

tq. $$L_{C'} < 6.51$$

Leçon II.4 (compression) – Étude de cas

Vous souhaitez envoyer un message, qui,

- considéré bit à bit, a une entropie de 0.9 bit,
- et considéré octet par octet, une entropie de 5.51 bit.

Sachant que ce message (d'origine, complet) a une taille de 10 Ko,
quelle taille pouvez-vous espérer après compression (sans autre information que celles fournies ici) ?

entre $10 \times \frac{5.51}{8}$ et $10 \times \frac{6.51}{8}$

(soit entre environ 7 et 8 Ko)

Leçon II.4 (compression) – Étude de cas

Considérons la séquence $X =$« HUBERT QUEL HURLUBERLU » (sans les espaces).

1. entropie ?
2. code de Shannon-Fano ?
3. code de Hufmann ?
4. conclure

H	U	B	E	R	T	Q	L	
2	5	2	3	3	1	1	3	→ 20

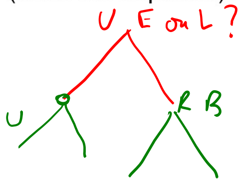
!! tri

U	E	R	L	H	B	T	Q
5	3	3	3	2	2	1	1

Leçon II.4 (compression) – Étude de cas

Considérons la séquence $X =$« HUBERT QUEL HURLUBERLU » (sans les espaces).

1. entropie ?
2. code de Shannon-Fano ?
3. code de Hufmann ?
4. conclure



U	E	L	R	B	H	Q	T	
5	3	3	3	2	2	1	1	20

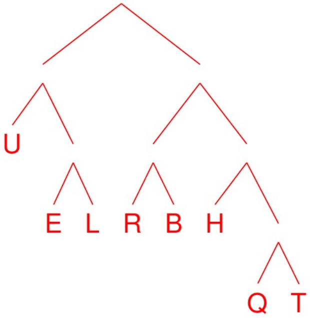
Leçon II.4 (compression) – Étude de cas

U	E	L	R	B	H	Q	T	
5	3	3	3	2	2	1	1	20

$$\begin{aligned}
H(X) &= \frac{1}{4}\log 4 + 3\times\frac{3}{20}\log\frac{20}{3} + 2\times\frac{1}{10}\log 10 + 2\times\frac{1}{20}\log 20 \\
&= \frac{1}{2} + \frac{9}{20}(2+\log 5-\log 3) + \frac{3}{10}(1+\log 5) + \frac{1}{10} \\
&= \frac{9}{5} + \frac{3}{4}\log 5 - \frac{9}{20}\log 3 \\
&\simeq 2.83 \text{ bit}
\end{aligned}$$

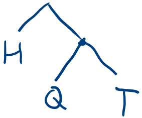
Leçon II.4 (compression) – Étude de cas

U	E	L	R	B	H	Q	T	
5	3	3	3	2	2	1	1	20



U	E	L	R	B	H	Q	T
0.25	0.15	0.15	0.15	0.1	0.1	0.05	0.05
2	3	3	3	3	3	4	4
0.5	0.45	0.45	0.45	0.3	0.3	0.2	0.2

$$L(C_{\text{S-F}}) = 2.85$$



au choix

2 à 2

U	E	L	R	B	H	Q	T	
5	3	3	3	2	2	1	1	20

4

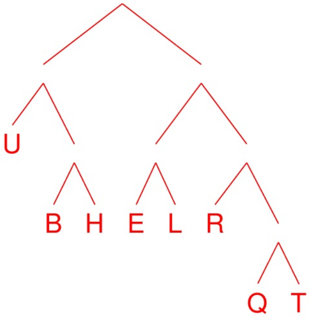
Huffman :

(1) les deux moins probables

U	E	L	R	B	H	Q	T
0.25	0.15	0.15	0.15	0.1	0.1	0.05	0.05
2	3	3	3	3	3	4	4
0.5	0.45	0.45	0.45	0.3	0.3	0.2	0.2

(2) idem

((3) etc...)



$$L(C_{\text{H}}) = 2.85$$

On a donc ici (mais ce n'est pas forcément toujours le cas) :

$$H(X) < L(C_{\text{H}}) = L(C_{\text{S-F}})$$