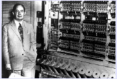


Information, Calcul et Communication

Compléments de cours

J.-C. Chappelier

EPFL

Leçon II.3 – Examen final 2018 Q1.1

$H(X) = -\sum_{x_i} p_i \log p_i$

x :	R	B	V	O
p_i :	$\frac{12}{24}$	$\frac{8}{24}$	$\frac{2}{24}$	$\frac{2}{24}$

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 $a + b\log_2(3)$.

$\log_2 24 - \frac{1}{24}\left(12\log_2 12 + 8\log_2 8 + 4\log_2 2\right)$

$3 + \log 3 - \frac{1}{2}(2 + \log 3) - \frac{1}{3}3 - \frac{1}{6}$

a	b
5/6	1/2
4/3	1
2/3	1/2
9/2	1/2

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

0 0 1 0 1

0 1 0 1 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]** $N-1$

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

$\lg(n+m) - \frac{1}{n+m}\left(n \lg n + m \log m\right)$

alphabet

	F	G
p_i:	$\frac{n}{n+m} = p$	$\frac{m}{n+m} \to 1-p$

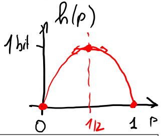
Quelle est l'entropie (telle que définie en cours) d'un mot X constitué de n fois la lettre 'F' et m fois la lettre 'G' : « $\underbrace{\texttt{F......F}}_{n}\underbrace{\texttt{G......G}}_{m}$ » ? $\to n+m$

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

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

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

$$= h(p)$$



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 :

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

Leçon II.3 – Examen 2 2016 Q6

Considérons une séquence de caractères de la forme « $AB\underset{\leftarrow 6 \rightarrow}{******}$ », = " X " 8 lettres

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 ?

maximale : $X_1 = AB\boxed{CDABCD}$ ✓ $\longrightarrow H(X_1) = 2$ bits

minimale : $X_3 = \cancel{ABABABAB}$ $\qquad \underbrace{H(X_2)}_{h\left(\frac{1}{8}\right)} < \underbrace{H(X_3)}_{1\,\text{bit}}$

$H(X_4) = H(X_2)$ $\qquad X_2 = AB\boxed{BBBBBB}$

$X_4 = ABAAAAAA$

Leçon II.3 – Examen 2 2016 Q6

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

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 :

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

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

$\exists\, C$:

$L(C) - 1 < H(X) \leq L(C)$

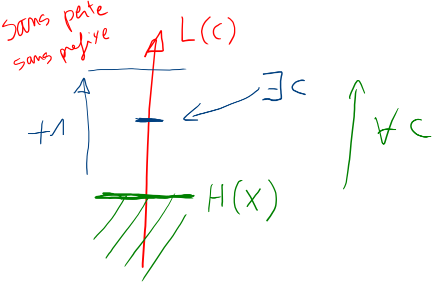
$\forall X$ source (alphabet, probas)

① $\forall C$ code sans préfixe sans perte de X $\quad H(X) \leq L(C(X))$

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

p.ex. S-F

Huff


Sans perte
sans préfixe
L(C)
+1
∃C
∀C
H(X)

Code : valeur $\longrightarrow$ valeur'

Leçon II.4 (compression) – Étude de cas

$$8$$

$$\boxed{0\,1\,0\,1\,1\,1\,0\,1}\;00$$

$1\ 0 \qquad C \qquad 0_{10} \qquad L(C)$

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. ← X

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) ?

$$H(X) \leq L(C) \quad \overset{?}{\leq} H(X)+1$$

5.51 bit

$$\text{gain} \quad \frac{8}{L(C)} \qquad \frac{L(C)}{8} \times 10\,\text{Ko}$$

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).

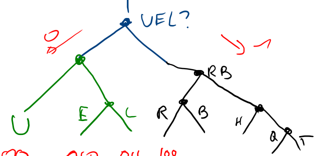
1. entropie ?
2. code de Shannon-Fano ?
3. code de Hufmann ?
4. conclure

S-F

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

2

3



	00	010	011	100
L_i	2	3	3	
p_i	5/20	3/20		

Huf

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

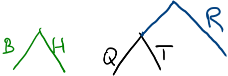
$1+1 = 2$

1 !

2

4

3 : 2 3



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

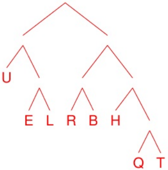
Leçon II.4 (compression) – Étude de cas

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

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

Leçon II.4 (compression) – Étude de cas

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

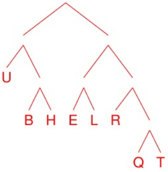


U	E	L	R	B	H	Q	T
0.25	0.15	0.15	0.15	0.1	0.1	0.05	0.05
2	3	3	3	3	3	4	4
0.5	0.45	0.45	0.45	0.3	0.3	0.2	0.2

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

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

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

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

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