Série 3

Olivier Cloux

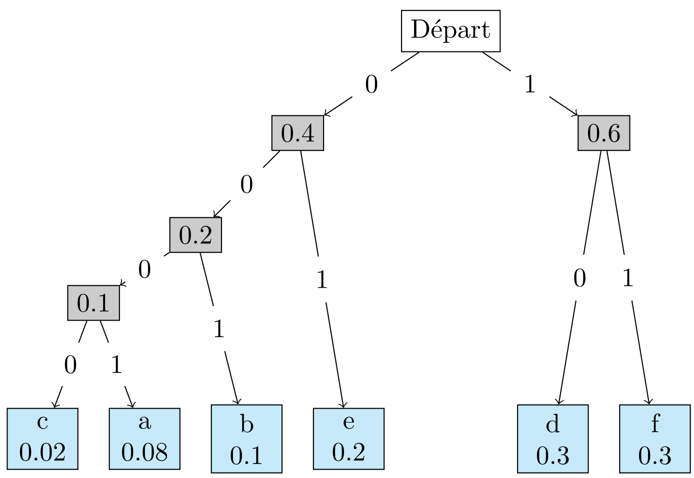
236079

Problème 3.1

1. Les longueurs des mots de code binaires de Shannon-Fano s'obtiennent avec la formule $\lceil \log_2 \left(\frac{1}{p_i}\right) \rceil$. En l'appliquant à chaque probabilité des symboles de source, nous obtenons :

Symbole de source	a	b	c	d	e	f
Probabilité	0.08	0.1	0.02	0.3	0.2	0.3
Longueur non-arrondie	3.64	3.32	5.64	1.73	2.32	1.73
Longueur arrondie	4	4	6	2	3	2

Un code de Huffman s'obtient en remplissant un arbre, selon la méthode vue en cours. L'arbre en question est :



Nous n'avons donc qu'à lire les positions afin de trouver le code associé :

Symbole de source	a	b	
Probabilité	0.08	0.1	0
Code de Huffman	0001	001	00

La longueur moyenne de ce code de Huffman est

$$0.08 \cdot 4 + 0.1 \cdot 3 + 0.02 \cdot 4 + 0.3 \cdot 2 + 0.2 \cdot 2 + 0.3 \cdot 2 = \boxed{2.3 = L(\Gamma_H)}$$

La longueur moyenne du code de Shannon-Fano est

$4 \cdot 0.08 + 4 \cdot 0.1 + 6 \cdot 0.02 + 2 \cdot 0.3 + 3 \cdot 0.2 + 2 \cdot 0.3 = \boxed{2.64 = L(\Gamma_{SF})}$

L'entropie de la source S, selon la méthode vue en cours, est

$0.08 \cdot \log_2\left(\frac{1}{0.08}\right) + 0.1 \cdot \log_2\left(\frac{1}{0.1}\right) + 0.02 \cdot \log_2\left(\frac{1}{0.02}\right) + 0.3 \cdot \log_2\left(\frac{1}{0.3}\right) + 0.2 \cdot \log_2\left(\frac{1}{0.2}\right) + 0.3 \cdot \log_2\left(\frac{1}{0.3}\right) = \boxed{2.24 = H(S)}$

Ainsi, l'entropie est bien la plus petite valeur, et le code de Huffman est bien optimal (en tout cas par rapport au code de Shannon-Fano).

La seconde inégalité de l'entropie est

$$\frac{H(S)}{\log_2(D)} \leq L(\Gamma_{SF}) < \frac{H(S)}{\log_2(D)} + 1$$

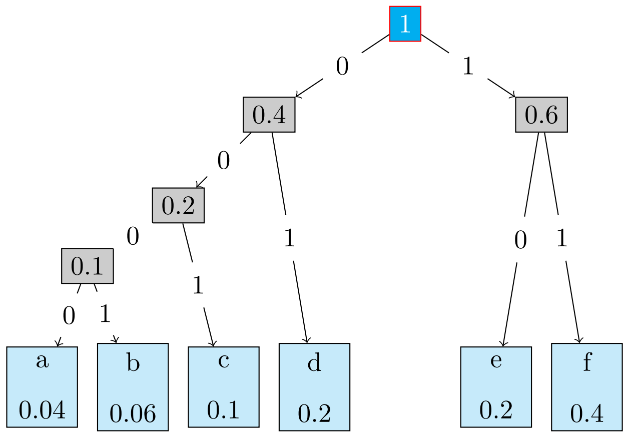
et elle est bien respectée pour le code de Shannon-Fano car

$$2.24 \leq 2.64 < 3.24 = 2.24 + 1$$

de même que pour le code de Huffman (qui est optimal, donc meilleur ou égal à Shannon-Fano), car

$$2.24 \leq 2.3 < 3.24$$

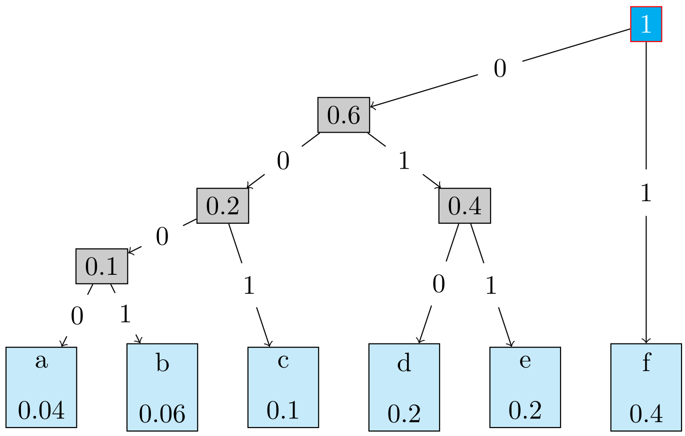
2. <u>Premier arbre de Huffman (H_1)</u>



Avec le code associé

Symbole de source	a	b	c	d	e	f
Probabilité	0.04	0.06	0.1	0.2	0.2	0.4
Γ_{H_1}	0000	0001	001	01	10	11

<u>Second arbre de Huffman (H_2):</u>



Avec le code associé

Symbole de source	a	b	c	d	e	f
Probabilité	0.04	0.06	0.1	0.2	0.2	0.4
Γ_{H_2}	0000	0001	001	010	011	1

$$L(\Gamma_{H_1}) = 2 \cdot 0.4 + 2 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.1 + 4 \cdot 0.06 + 4 \cdot 0.04 = 2.3$$

$$L(\Gamma_{H_2}) = 1 \cdot 0.4 + 3 \cdot 0.2 + 3 \cdot 0.2 + 3 \cdot 0.1 + 4 \cdot 0.06 + 4 \cdot 0.04 = 2.3$$

3. Le théorème 3.4 du code de Huffman nous dit que

   [...]pour tout autre code binaire à décodage unique Γ on a

$L(\Gamma_H) \leq L(\Gamma)$

Ce qui signifie que

$L(\Gamma_{H_1}) \leq L(\Gamma)$

mais aussi que

$L(\Gamma_{H_2}) \leq L(\Gamma)$

Donc que $(\Gamma_{H_2}) \leq (\Gamma_{H_1}) \leq (\Gamma_{H_2}) \leq (\Gamma)$

ce qui implique que $(\Gamma_{H_2}) = (\Gamma_{H_1})$

Problème 3.2

1. Nous pouvons utiliser la même formule qu'à l'exercice 3.1.1 ($\lceil \log_2 \left(\frac{1}{p_i}\right) \rceil$)

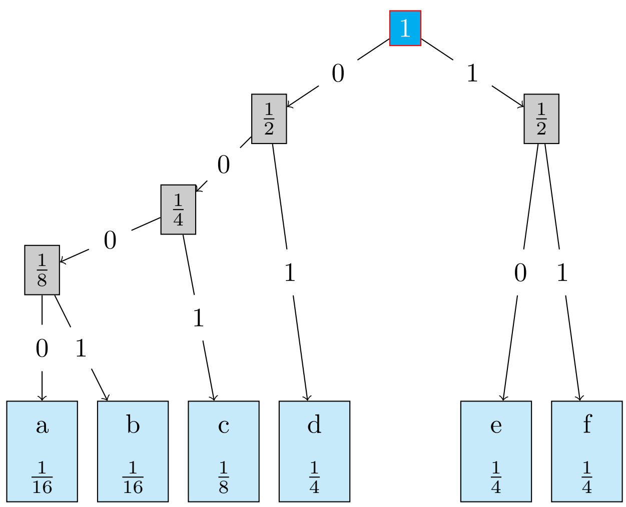
   Cela nous donne les longueurs suivantes

Symbole de source	a	b	c	d	e	f
Probabilité	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Longueur non-arrondie	4	4	3	2	2	2
Longueur arrondie	4	4	3	2	2	2

Pour une longueur moyenne

$L(\Gamma_{SF}) = \frac{4}{16} + \frac{4}{16} + \frac{3}{8} + \frac{2}{4} + \frac{2}{4} + \frac{2}{4} = \boxed{2.375 = L(\Gamma_{SF})}$

L'arbre de Huffman associé au code S est :



Avec le code qui en découle :

Symbole de source	a	b	c	d	e	f
Probabilité	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Γ_H	0000	0001	001	01	10	11

La longueur moyenne de ce code est :

$\frac{4}{16} + \frac{4}{16} + \frac{3}{8} + \frac{2}{4} + \frac{2}{4} + \frac{2}{4} = \boxed{2.375 = L(\Gamma_H)}$

L'entropie de la source S est

$2 \cdot \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + 3 \cdot \frac{1}{4}\log_2(4) = \boxed{2.375 = H(S)}$

Il est alors amusant de constater (nous le prouverons à la question suivante) que les 3 longueurs sont exactement les mêmes.

2. (a) Ce qui fait varier la longueur des mots de code de Shannon par rapport à l'entropie, c'est l'arrondi vers le haut. En effet, là où l'entropie multiplie la probabilité d'apparition par le log de l'inverse de cette probabilité ($\sum p(s)\log_2(\frac{1}{p(s)})$), la longueurs selon Shannon multiplie la probabilité d'apparition par <u>l'arrondi vers le haut</u> de ce même logarithme ; la longueur du mot de code est $\lceil \log_2(\frac{1}{p(s)}) \rceil$, et pour trouver la longueur moyenne nous multiplions les mots de code par leur probabilité d'apparition, ce qui nous donne $\sum p(s)\lceil \log_2(\frac{1}{p(s)}) \rceil$. Mais lorsque les $p(s)$ sont des puissances de 2, alors $\lceil \log_2(\frac{1}{p(s)}) \rceil = \log_2(\frac{1}{p(s)})$ (car $\log_2(2^n) = n$). La longueurs des mots de Shannon ont alors la même longueur (car même formule) que l'entropie.

   (b) Comme nous l'avons dit avant, $L(\Gamma_{SF}) = \sum p(s)\lceil \log_2(\frac{1}{p(s)}) \rceil = \sum p(s)\log_2(\frac{1}{p(s)}) = H(S)$ quand et seulement quand tous les $p(s)$ sont des puissances de 2, car l'arrondi n'a plus lieu (le $\log_2$ étant alors déjà entier). Ainsi, lorsque $L(\Gamma_{SF})$ devient plus grand que l'entropie, cela signifie qu'une longueur a été arrondie vers le haut (selon la définition), et donc qu'un des $p(s)$ n'est plus une puissance de 2.

<u>Problème 3.3</u>

1. $\begin{array}{c|c} A & B \\ \hline \frac{3}{4} & \frac{1}{4} \end{array}$. L'entropie est donc $\frac{3}{4}\log_2\left(\frac{4}{3}\right) + \frac{1}{4}\log_2(4) = \boxed{0.811 = H(S_0)}$

2. $H(S_1|S_0) = H(S_1|S_0 = A)p_{S_0}(A) + H(S_1|S_0 = B)p_{S_0}(B)$

   $H(S_1|S_0 = A) = \frac{1}{2}\log_2(2) + \frac{1}{2}\log_2(2) = 2\frac{1}{2}\log_2(2) = 1$

   $\big($car 2 issues possibles (P/F) avec chacun probabilité 1/2$\big)$

   $H(S_1|S_0 = B) = 1\log_2(1) = 0$

   $\big($car seulement F possible, avec probabilité 1$\big)$

   $p_{S_0}(A) = \frac{3}{4}, p_{S_0}(B) = \frac{1}{4} \to H(S_1|S_0) = \frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 0 = \boxed{\frac{3}{4} = H(S_1|S_0)}$

   $H(S_2|S_0) = H(S_2|S_0 = A)p_{S_0}(A) + H(S_2|S_0 = B)p_{S_0}(B)$

   $H(S_2|S_0 = A) = \frac{1}{2}\log_2(2) + \frac{1}{2}\log_2(2) = 2\frac{1}{2}\log_2(2) = 1$

   $H(S_2|S_0 = B) = 1\log_2(1) = 0$

   (mêmes justifications que pour $H(S_1|S_0)$)

$p_{S_0}(A) = \frac{3}{4}, p_{S_0}(B) = \frac{1}{4} \to H(S_2|S_0) = \frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 0 = \boxed{\frac{3}{4} = H(S_2|S_0)}$

$H(S|S_0) = H(S|S_0 = A)p_{S_0}(A) + H(S|S_0 = B)p_{S_0}(B)$

$p_{S_0}(A) = \frac{3}{4},\ p_{S_0}(B) = \frac{1}{4}$

$H(S|S_0 = B) = 1 \log_2(1) = 0$

(car seulement FF possible, avec probabilité 1)

$H(S|S_0 = A) = 4 \cdot \frac{1}{4} \log_2(4) = 2$

(car 4 issues possibles PF, FP, FF, PP, chacun avec probabilité $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$) $\to H(S|S_0) = \frac{3}{4} \cdot 2 + \frac{1}{4} \cdot 0 = \boxed{\frac{3}{2} = H(S|S_0)}$

Nous remarquons - à notre immense étonnement - que $H(S_2|S_0) = H(S_1|S_0)$. En effet, l'ordre des tirages n'a pas d'importance, une fois que l'on connaît S_0. Pile arrivera avec la même probabilité que face ($\frac{1}{2}$) si la pièce A est prise, et Face arrivera forcément si B est prise, les probabilités ne changent pas $\to \boxed{H(S_2|S_0) = \frac{3}{4} = H(S_1|S_0)}$

3. Pour S_1 nous avons la densité de probabilité suivante :

Outcome possible	Probabilité
P	3/8
F	5/8

En effet, $p(P) = p(AP) + P(BP) = \frac{3}{8} + 0$
$p(F) = p(AF) + P(BF) = \frac{3}{8} + \frac{1}{4} = \frac{5}{8}$

$H(S_1) = \frac{3}{8} \log_2(\frac{8}{3}) + \frac{5}{8} \log_2(\frac{8}{5}) = \boxed{0.954 = H(S_1)}$

Pour S_2, la densité de probabilité est

Outcome possible	Probabilité
P	3/8
F	5/8

(Mêmes explications que pour S_1). La densité de probabilité étant la même, l'entropie est donc la même aussi $\to \boxed{0.954 = H(S_2)}$

Pour S en revanche, la densité de probabilité change légèrement :

Outcome possible	Probabilité
PP	3/16
PF	3/16
FP	3/16
FF	7/16

$$\text{Car} \left\{ \begin{array}{l} p(PP) = p(APP) + p(BPP) = \frac{3}{16} + 0 = \frac{3}{16} \\ p(PF) = p(APP) + p(BPF) = \frac{3}{16} + 0 = \frac{3}{16} \\ p(FP) = p(AFP) + p(BFP) = \frac{3}{16} + 0 = \frac{3}{16} \\ p(PP) = p(APP) + p(BPP) = \frac{3}{16} + \frac{1}{4} = \frac{7}{16} \end{array} \right\}$$

Donc $H(S) = 3 \cdot \frac{3}{16} \log_2(\frac{16}{3}) + \frac{7}{16} \log_2(\frac{16}{7}) = \boxed{1.88 = H(S)}$. $H(S_1|S_0) = \frac{3}{4} \leq 0.954 = H(S_1)$

et $H(S_1, S_2|S_0) = \frac{3}{2} \leq 1.88 = H(S)$. Cela semble correcte, car ces inégalités répondent au théorème 4.2(Conditionner réduit l'entropie). En effet, savoir S_0 nous apporte une information supplémentaire (caractérisée par la différence entre les deux termes d'une inégalité) ; ayant plus d'information, l'entropie s'en trouve donc réduite.

4. $p_{S_2|S_1}(s_2|s_1)$ = "probabilité que $S_2 = s_2$ sachant que $S_1 = s_1$" $= \frac{p(s_1,s_2)}{p(s_1)}$ Nous pouvons donc analyser les cas un par un[1] :

   - $p_{S_2|S_1}(P|P) = \frac{1}{2}$ Une fois que le premier tirage est fait et que pile est sorti, on sait que la pièce est A. La probabilité d'avoir pile est donc de $\frac{1}{2}$
   - $p_{S_2|S_1}(F|P) = \frac{1}{2}$, pour la même raison qu'au-dessus. Pile est sorti, donc nous savons que la pièce est A, donc une chance sur deux d'avoir pile, et une demi pour face.
   - $p_{S_2|S_1}(P|F) = \frac{p_{S_1,S_2}(FP)}{p_{S_1}(F)} = \frac{\frac{3}{16}}{\frac{5}{8}} = \frac{3}{10}$
   - $p_{S_2|S_1}(F|F) = \frac{p_{S_1,S_2}(FF)}{p_{S_1}(F)} = \frac{\frac{7}{10}}{\frac{5}{8}} = \frac{7}{10}$

5. $H(S_2|S_1) = \sum\limits_{s_1 \in A_1} H(S_2|S_1 = s_1)p_{S_1}(s_1)$
   $\rightarrow = H(S_2|S_1 = P)p_{S_1}(P) + H(S_2|S_1 = F)p_{S_1}(F)$
   $p_{S_1}(P) = \frac{3}{8},\ p_{S_1}(F) = \frac{5}{8}$
   $H(S_2|S_1 = P) = 2 \cdot \frac{1}{2}\log_2(2) = 1$
   $H(S_2|S_1 = F) = \frac{3}{10}\log_2(\frac{10}{3}) + \frac{7}{10}\log_2(\frac{10}{7}) \simeq 0.8813$
   $\rightarrow H(S_2|S_1) = \frac{5}{8} \cdot 0.8813 + \frac{3}{8} \cdot 1 \simeq \boxed{0.926 = H(S_2|S_1)}$
   Cette entropie est plus faible que celle de S_2 (ce qui répond au théorème 4.2), car une fois que l'on sait S_1, nous sommes dirigés vers une solution : Si face apparaît, nous serons tentés de proposer face au tirage suivant. Cette tentation est caractérisée par une entropie inférieure.

6. La règle de l'enchaînement pour l'entropie semble être :
   $H(S_2|S_1) = H(S_2, S_1) - H(S_1) = H(S) - H(S_1) = 1.88 - 0.954 = 0.926$. La valeur de cette entropie est donc vérifiée.

[1] c.f. partie 3 pour les probabilités

7. La densité de probabilité est (dans l'ordre S_0, S_1, S_2)

Outcome possible	probabilité
APP	3/16
APF	3/16
AFP	3/16
AFF	3/16
BPP	0
BPF	0
BFP	0
BFF	1/4

Donc l'entropie de cette source est : $4 \cdot \frac{3}{16}\log_2(\frac{16}{3}) + \frac{1}{4}\log_2(4) = \boxed{2.311 = H(S_0, S_1, S_2)}$

8. $H(S_2|S_1, S_0) = H(S_2|S_1, S_0 = PA) \cdot p_{S_1,S_0}(PA) + H(S_2|S_1, S_0 = FA) \cdot p_{S_1,S_0}(FA) + H(S_2|S_1, S_0 = PB) \cdot p_{S_1,S_0}(PB) + H(S_2|S_1, S_0 = FB) \cdot p_{S_1,S_0}(FB)$

Outcome	proba
$p_{S_1,S_0}(FB)$	$\frac{1}{4}$
$p_{S_1,S_0}(PB)$	0
$p_{S_1,S_0}(FA)$	$\frac{3}{8}$
$p_{S_1,S_0}(PA)$	$\frac{3}{8}$

$H(S_2|S_1, S_0 = PA) = H(S_2|S_1, S_0 = FA) = 2 \cdot \frac{1}{2}\log_2(2) = 1$

$H(S_2|S_1, S_0 = PA) = 0$ (cas impossible)

$H(S_2|S_1, S_0 = FB) = 1\log_2(1) = 0$ (source déterministe).

$\rightarrow H(S_2|S_1, S_0) = 2 \cdot \frac{3}{8} \cdot 1 = \boxed{\frac{3}{4} = H(S_2|S_1, S_0)}$.

Nous voyons que $H(S_2|S_1, S_0) = H(S_2|S_0)$. En effet, savoir S_1 en plus ne nous donnera aucune information en plus ; Si la pièce est A, alors Pile et Face arriveront toujours avec la même probabilité (jets indépendants), et si la pièce est B alors Face sortira tout le temps $\rightarrow$ connaître le premier tirage (si l'on sait quelle pièce est utilisée) ne nous fournit aucune information supplémentaire sur S_2.

9. $H(S_0|S_1, S_2) = H(S_0|S) = H(S_0, S) - H(S) = H(S_0, S_1, S_2) - H(S_1, S_2) = 2.311 - 1.88 = \boxed{0.431 = H(S_0|S_1, S_2)}$