

## Série 4

### Problème 4.1

Notation :  $S^n = (S_1, S_2, S_3, \dots, s_n)$ , P = pile, F = face

$$1. \quad \boxed{p_{S_0}(A) = 2 \cdot \frac{1}{4} = \frac{1}{2}}$$

$$\boxed{p_{S_0}(B) = p_{S_0}(C) = \frac{1}{4}}$$

$$H(S_0) = \frac{1}{2} \log_2(2) + \frac{1}{4} \log_2(4) \cdot 2 = \boxed{\frac{3}{2} = H(S_0)}$$

2. Nous devons remarquer 3 cas distincts :

$$S^n = (F, F, F, F, F, \dots, F)$$

$$S^n = (P, P, P, P, P, \dots, P)$$

$$S^n = (P, F, P, P, \dots) \text{ (tous les autres cas, qui mélangent P et F)}$$

Il y a deux manières d'obtenir le premier cas : tirer la pièce B (donc que des F seront engendrés après), ou tirer une pièce A et tirer, par hasard, que des F. Tirer A et que des F est difficile à avoir : à  $n$  tirages, il y a  $2^n$  cas (dont un seul nous intéresse), alors que avec B, la probabilité d'avoir F est de 1 (en fait  $1^n$  au  $n^{\text{ème}}$  tirage). Nous trouvons donc que

$$p_{S^n}(FFF\dots) = \frac{1}{4} \cdot 1^n + \frac{1}{2} \frac{1}{2^n} = \boxed{p_{S^n}(FFF\dots) \frac{1}{4} + \frac{1}{2^{n+1}} = \alpha_n}$$

En travaillant un peu  $\alpha_n$ , nous trouvons

$$\alpha_n = \frac{1}{4} + \frac{1}{2^{n+1}} = \frac{2^{n-1}}{2^{n+1}} + \frac{1}{2^{n+1}} = \frac{2^{n-1}+1}{2^{n+1}}$$

Le cas “que des P” est le même que le premier cas (à la différence que l'on souhaite tirer la pièce C ou une A et que des P). La probabilité ne change pas, et  $\boxed{\text{nous obtenons donc aussi } \alpha_n \text{ pour } p_{S^n}(PPP\dots)}$ .

Quant aux autres cas, la seule possibilité est de tirer une pièce A et

$$\text{d'ensuite tirer la séquence exacte. Donc } \boxed{p_{S^n}(\text{reste}) = \frac{1}{2} \frac{1}{2^n} = \frac{1}{2^{n+1}} = \alpha_n - \frac{1}{4} = \beta_n}$$

3. Notre source  $S^n$  produit une suite de  $n$  symboles venant chacun de l'alphabet  $D = \{P, F\}$ . L'alphabet de  $S^n$  est  $D^n$ . Selon la définition, cette source est donc un bloc. Et, comme nous ne précisons pas le nombre de tirages de la pièce (mais nous le supposons très grand, voire infini), le bloc est infini, donc la source est une **Source étendue** #
4. Une source est stationnaire si chaque bloc possible a la même probabilité d'apparition. Comme nous venons de le voir, deux blocs - par exemple (FFFFFF....) et (PFPFPFPF....) - n'ont pas du tout la même probabilité d'apparition. Notre source, donc, n'est pas stationnaire.
5. A  $n$  tirs, il y a  $2^n$  issues possibles :
  - 2 cas (que P ou que F) à probabilité  $\alpha_n$ .
  - $2^n - 2$  cas (les autres) à probabilité  $\frac{1}{2^{n+1}}$

L'entropie est donc :

$$\begin{aligned}
 H(S^n) &= \alpha_n \log_2\left(\frac{1}{\alpha_n}\right) \cdot 2 + \frac{1}{2^{n+1}} \log_2(2^{n+1}) \cdot (2^n - 2) \\
 &= 2\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) + \frac{(n+1)(2^n-2)}{2^{n+1}} \\
 &= 2\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) + (n+1) \frac{2^n-2}{2^{n+1}} \\
 &= 2\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) + (n+1) \left(\frac{2^n}{2^{n+1}} - \frac{2}{2^{n+1}}\right) \\
 &= 2\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) + (n+1) \left(\frac{1}{2} - \frac{1}{2^n}\right) \\
 &= 2\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) + \frac{n+1}{2} - \frac{n+1}{2^n} = H(S^n) = \text{entropie d'un symbole}
 \end{aligned}$$

6. Nous savons que  $H(S_b|S_a) = H(S_a, S_b) - H(S_a)$ . Si nous posons ensuite  $S_n = S_b$  et  $S^{n-1} = S_a$ , nous pouvons poser :
 
$$H(S_n|S_1, S_2, \dots, S_{n-1}) = H(S_1, S_2, \dots, S_{n-1}, S_n) - H(S_1, \dots, S_{n-2}, S_{n-1})$$
 Calculer  $H(S_n|S^{n-1})$  devient alors plus facile, car nous connaissons  $H(S^n)$ , et par extension  $H(S^{n-1})$  (en remplaçant  $n$  par  $n-1$  dans notre calcul de l'entropie, précédemment)

Ainsi

$$\begin{aligned}
H(S_n|S^{n-1}) &= H(S^n) - H(S^{n-1}) \\
&= 2\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) + \frac{n+1}{2} - \frac{n+1}{2^n} - \left(2\alpha_{n-1} \log_2\left(\frac{1}{\alpha_{n-1}}\right) + \frac{(n-1)+1}{2} - \frac{(n-1)+1}{2^{n-1}}\right) \\
&= 2\left(\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) - \alpha_{n-1} \log_2\left(\frac{1}{\alpha_{n-1}}\right)\right) + \frac{n+1}{2} - \frac{n}{2} - \frac{n+1}{2^n} + \frac{n}{2^{n-1}} \\
&= 2\left(\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) - \alpha_{n-1} \log_2\left(\frac{1}{\alpha_{n-1}}\right)\right) + \frac{1}{2} - \frac{n+1}{2^n} + \frac{2n}{2^n} \\
&= 2\left(\alpha_n \log_2\left(\frac{1}{\alpha_n}\right) - \alpha_{n-1} \log_2\left(\frac{1}{\alpha_{n-1}}\right)\right) + \frac{1}{2} + \frac{n-1}{2^n}
\end{aligned}$$

Depuis là, nous pouvons commencer à analyser les cas numériques :

$$\begin{aligned}
n=10 \quad \alpha_n &= \frac{2^9+1}{2^{11}}, \quad \alpha_{n-1} = \frac{2^8+1}{2^{10}} \\
\rightarrow H(S_n|S^{n-1}) &= 2\left(\frac{2^9+1}{2^{11}} \log_2\left(\frac{2^{11}}{2^9+1}\right) - \frac{2^8+1}{2^{10}} \log_2\left(\frac{2^{10}}{2^8+1}\right)\right) + \frac{1}{2} + \frac{10-1}{2^{10}} \\
&= \boxed{0.50824} \\
n=15 \quad \alpha_n &= \frac{2^{14}+1}{2^{16}}, \quad \alpha_{n-1} = \frac{2^{13}+1}{2^{15}} \\
\rightarrow H(S_n|S^{n-1}) &= 2\left(\frac{2^{14}+1}{2^{16}} \log_2\left(\frac{2^{16}}{2^{14}+1}\right) - \frac{2^{13}+1}{2^{15}} \log_2\left(\frac{2^{15}}{2^{13}+1}\right)\right) + \frac{1}{2} + \frac{15-1}{2^{15}} \\
&= \boxed{0.50041} \\
n=20 \quad \alpha_n &= \frac{2^{19}+1}{2^{21}}, \quad \alpha_{n-1} = \frac{2^{18}+1}{2^{20}} \\
\rightarrow H(S_n|S^{n-1}) &= 2\left(\frac{2^{19}+1}{2^{21}} \log_2\left(\frac{2^{21}}{2^{19}+1}\right) - \frac{2^{18}+1}{2^{20}} \log_2\left(\frac{2^{20}}{2^{18}+1}\right)\right) + \frac{1}{2} + \frac{20-1}{2^{20}} \\
&= \boxed{0.50002}
\end{aligned}$$

7. Rappelons, pour les ingénus qui auraient oublié une telle définition, ce qu'est une source régulière :

**Une source étendue  $S$  est dite *régulière* si les deux limites**

- $H(S) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} H(S^n)$  et
- $H^*(S) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} H(S_n|S^{n-1})$

**existent et sont finies.** Nous pouvons donc calculer les limites. Commençons  $H(S)$  :

$$\begin{aligned}
\lim_{n \rightarrow \infty} \alpha_n &\simeq \lim_{n \rightarrow \infty} \frac{2^{n-1}}{2^{n+1}} = \frac{1}{4} \\
\lim_{n \rightarrow \infty} \alpha_{n-1} &= \lim_{n \rightarrow \infty} \alpha_n = \frac{1}{4} \\
\lim_{n \rightarrow \infty} \frac{n+1}{2^n} &\simeq \lim_{n \rightarrow \infty} \frac{n}{2^n} = 0 \\
\lim_{n \rightarrow \infty} \log_2\left(\frac{1}{\alpha_n}\right) &= \log_2\left(\lim_{n \rightarrow \infty} \frac{1}{\alpha_n}\right) = \log_2(4) = 2
\end{aligned}$$

Donc  $\lim_{n \rightarrow \infty} H(S^n) = \infty$ .

mais :

$$\lim_{n \rightarrow \infty} \frac{n+1}{2} = \infty$$

Donc la source ne peut pas être régulière, car  $H(S)$  n'existe pas. Calculons

quand même l'entropie par symbole :

Nous pouvons reprendre les valeurs des limites calculées précédemment.

Cela nous donne donc :

$$\begin{aligned}\lim_{n \rightarrow \infty} H(S_n | S^{n-1}) &= \lim_{n \rightarrow \infty} 2 \left( \alpha_n \log_2 \left( \frac{1}{\alpha_n} \right) - \alpha_{n-1} \log_2 \left( \frac{1}{\alpha_{n-1}} \right) \right) + \frac{1}{2} + \frac{n-1}{2^n} \\ &= 2 \left( \frac{1}{4} \log_2(4) - \frac{1}{4} \log_2(4) \right) + \frac{1}{2} + 0 \\ &= 2 \cdot (0) + \frac{1}{2} = \boxed{\frac{1}{2}}\end{aligned}$$

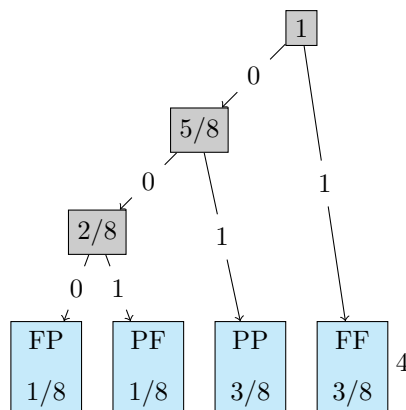
8. Calculons cette limite :

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{H(S^n)}{n} &= \lim_{n \rightarrow \infty} \frac{2\alpha_n \log_2 \left( \frac{1}{\alpha_n} \right) + \frac{n+1}{2} - \frac{n+1}{2^n}}{n} \\ &= \lim_{n \rightarrow \infty} \frac{2\alpha_n \log_2 \left( \frac{1}{\alpha_n} \right) + \frac{n+1}{2} - \frac{n+1}{2^n}}{\lim_{n \rightarrow \infty} 2\alpha_n \log_2 \left( \frac{1}{\alpha_n} \right) + \frac{n+1}{2} - \frac{n+1}{2^n}} \\ &= \lim_{n \rightarrow \infty} \frac{2\alpha_n \log_2 \left( \frac{1}{\alpha_n} \right) - \frac{n+1}{2^n}}{n} + \lim_{n \rightarrow \infty} \frac{\frac{n+1}{2}}{n} \quad \text{Nous remarquons} \\ &= \text{c.f. plus haut } \frac{2 \cdot \frac{1}{4} \log_2(4) - 0}{\infty} + \lim_{n \rightarrow \infty} \frac{n+1}{2n} \\ &= 0 + \frac{1}{2} \\ &= \boxed{\frac{1}{2}}\end{aligned}$$

que cette limite est la même que  $H^*(S)$ . En effet,  $H^*(S)$  nous dira combien de d'information supplémentaire apportera le dernier tirage, en connaissant les précédents. La limite que nous venons de calculer nous donne le nombre de bits pour un tirage supplémentaire (entropie totale divisée par le nombre de tirages). Cela signifie que l'entropie supplémentaire d'un tirage connaissant les précédents est la même que l'entropie de chaque tirage (à l'infini). De plus, nous savons par le théorème 5.2 que cette égalité doit être respectée.

#### **Problème 4.2**

1.
  - Probabilité de  $PP$  et de  $FF$  :  $\frac{1}{4} + \frac{1}{2} \cdot \frac{1}{2^2} = \frac{3}{8}$
  - Probabilité de  $PF$  et de  $FP$  :  $\frac{1}{2} \cdot \frac{1}{2^2} = \frac{1}{8}$



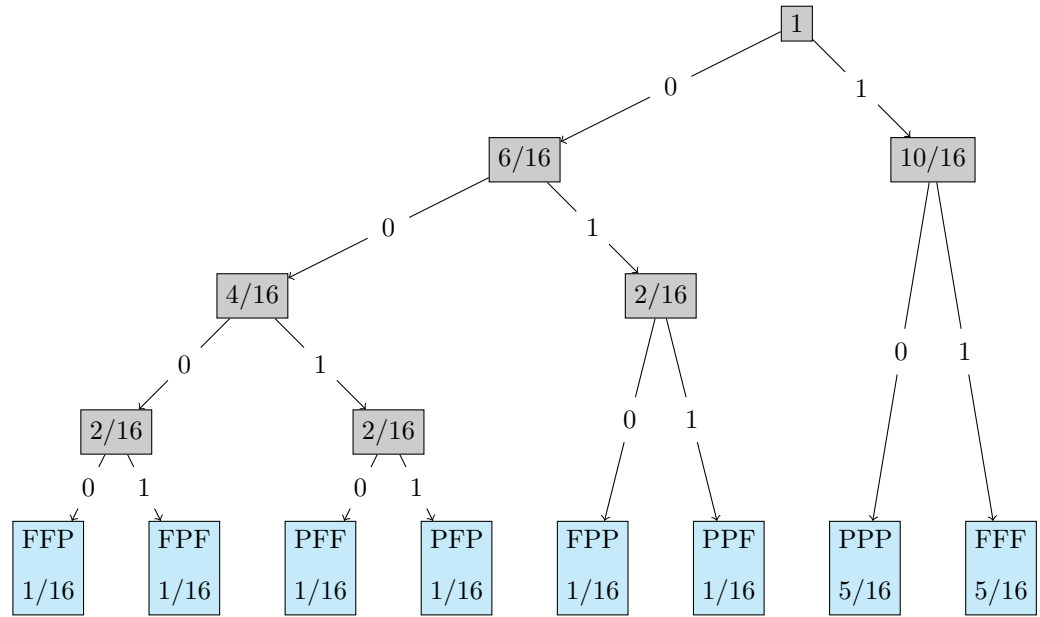
Symbole de source	PF	FP	PP	FF
Probabilité	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$
$\Gamma_{H2}$	000	001	01	1

$$L(\Gamma_{H2}) = 2 \cdot \frac{1}{8} 3 + \frac{3}{8} (1 + 2) = \frac{20}{8} = \boxed{1.875 = L(\Gamma_{H2})}$$

$$H(S^2) = 2 \cdot \frac{1}{8} \log_2(8) + 2 \cdot \frac{3}{8} \log_2\left(\frac{8}{3}\right) = \boxed{1.811 = H(S^2)}$$

Comme nous nous y attendions, la longueur moyenne du code de Huffman est plus longue que l'entropie. De plus, nous pouvons voir que  $\frac{L(\Gamma_{H2})}{2} = \frac{1.875}{2} = 0.9375 > H^*(S)$ . En effet,  $H^*(S)$  est un encodage par blocs, ce qui le rend plus efficace

2. • Probabilité de  $PPP$  et de  $FFF$  :  $\frac{1}{4} + \frac{1}{2} \frac{1}{2^4} = \frac{5}{16}$
- Probabilité de  $PPF$ ,  $PFF$ ,  $FFP$ ,  $PFP$ ,  $FPP$ ,  $FPP$  :  $\frac{1}{2} \frac{1}{2^3} = \frac{1}{16}$



Symbole de source	FFP	FPF	PFF	PFP	FPP	PPF	PPP	FFF
Probabilité	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{5}{16}$
$\Gamma_{H3}$	0000	0001	0010	0011	010	011	10	11

$$L(\Gamma_{H3}) = 4 \cdot \frac{1}{16} \times 4 + 3 \cdot \frac{1}{16} \times 2 + 2 \cdot \frac{5}{16} \times 2 = \boxed{2.625 = L(\Gamma_{H3})}$$

$$H(S^3) = \frac{1}{16} \log_2(16) \times 6 + \frac{5}{16} \log_2\left(\frac{16}{5}\right) \times 2 = \boxed{2.5488 = H(S^3)}$$

La différence entre  $L(\Gamma_{H3})$  et  $H(S^3)$ , avec l'entropie plus faible que la

Cependant,  $\frac{L(\Gamma_{H3})}{3} = \frac{2.625}{3} = 0.875 > 0.5 = H^*(S)$ . En effet,  $H^*(S)$  est un encodage par blocs, ce qui le rend plus efficace. Finalement, nous pouvons remarquer que, par rapport à  $\Gamma_{H3}$ , la longueur moyenne par bloc semble se rapprocher de  $H^*(S)$  ;

$$\frac{1}{4} + \frac{1}{2^{1001}} = \frac{2^{999}+1}{2^{1001}} 0.2500$$

000  
000  
000  
18092516094394950447723619085848085457231858540123108571698979834554878878172

(donc environ 0.25)

Il s'agit ensuite d'utiliser l'algorithme de Shannon pour cela : rappelons-nous que la longueur des mots de code, selon Shannon-Fano, est de  $\lceil \log_2(\frac{1}{p_i}) \rceil$ .

Pour les deux autres (uniquement pile et ou face), ils auront une longueur de

Cela nous donne une longueur moyenne de

$$\begin{aligned} L(\Gamma_{SF}) &= 1001 \frac{1}{2^{1001}} \times (2^{1000} - 2) + 2(\frac{1}{4} + \frac{1}{2^{1001}}) \times 2 \simeq 501.500000 \\ H(S^{1000}) &= 2^{\frac{999}{2^{1001}}+1} \log_2(\frac{2^{1001}}{999+1}) + \frac{1000+1}{2} - \frac{1000+1}{2^{1000}} \simeq 501.500000 \end{aligned}$$

Avec un nombre de tirages très grand, la probabilité de “que pile” et de

“que face” tend vers  $1/4$  (donc tirer une pièce truquée, car tirer 1000x de suite la même chose avec une pièce pas truquée est extrêmement faible), donc “que pile ou que face” tend vers  $1/2$ . De plus, l’entropie de Shannon va tendre vers  $n/2$ . Si nous cherchons la longueur moyenne par symbole, nous divisons encore cela par  $n$ , donc vers  $\frac{1/2+n/2}{n}$ , donc vers  $\frac{1}{2}$ . A l’infini, tend donc vers l’efficacité de l’entropie.

4. Pour les  $2^{1000} - 2$  cas banals, nous avons vu qu’ils auront une longueur encodée de 1001 bits. Avec 1000 lancés, le bloc  $(S_1, S_2, \dots, S_{1000})$  est de longueur 1000. Il suffit donc de changer chaque résultat de chaque lancé en un chiffre (par exemple Pile devient 1, Face devient 0), et d’ajouter un 0 avant (pour respecter la longueur de Shannon, et éviter les préfixes avec les 2 cas particuliers). De cette manière, comme chaque bloc  $(S_1, \dots, S_{1000})$  est unique, son encodage le sera aussi.

Les deux cas particuliers seront juste encodés sur 2 bits : le premier serait 1 (pour éviter les préfixes avec les cas banals), et le second serait 1 pour Pile et 0 pour Face.

La longueur moyenne de notre code est de 501.5. Transformer chaque tirage en 1 ou 0 selon la méthode proposée donnerait un code de longueur 1000. Le rapport de compression est donc  $\frac{501.5}{1000} = 0.5015$ , ce qui est notre rapport de compression. Notre méthode de codage nous permet donc de compresser considérablement le code.