Série 12: Solutions

1 L'entropie d'un jeu d'échecs

a) Voici un exemple de stratégie optimale (il y en a plusieurs équivalentes):

Q1: Est-ce une case vide? Si oui, on sait que la case est vide (avec 1 question posée). Si non, on continue:

Q2: La pièce est-elle noire? Q3: La pièce est-elle un pion?

Avec ces 2 questions supplémentaires, si la réponse à la question 3 est oui, on sait que la case contient un pion (blanc ou noir), avec donc 3 questions posées en tout. Si non, on continue:

Q4: Est-ce que la pièce est un cavalier ou un fou?

Si oui: Q5: Est-ce que la pièce est un cavalier? Avec les réponses à ces 2 dernières questions, on peut identfier la pièce (donc avec en 5 questions en tout).

Si non: Q5: Est-ce que la pièce est une tour? Si oui, on peut à nouveau identifier la pièce avec 5 questions en tout. Sinon, on doit encore poser une question: Q6: est-ce que la pièce est une dame? Et on obtient donc dans ce cas la réponse en 6 questions.

Au total, en tenant compte des probabilités d'apparition de chaque pièce sur l'échiquier, on trouve que le nombre moyen de questions à poser est:

$$\frac{32}{64} \times 1 + \frac{16}{64} \times 3 + \frac{12}{64} \times 5 + \frac{4}{64} \times 6 = \frac{1}{2} + \frac{3}{4} + \frac{15}{16} + \frac{3}{8} = \frac{41}{16} = 2.5625$$

b) Dans le diagramme de gauche, le nombre et la nature des pièces n'a pas changé par rapport à la situation de départ, et donc l'entropie reste la même, de la même façon que l'ordre des lettres dans une séquence de lettres n'influence pas l'entropie de la séquence: seules les probabilités d'apparition comptent dans le calcul de l'entropie.

Dans le diagramme de droite, un pion blanc a disparu. On pourrait faire ici un calcul long et compliqué pour recalculer l'entropie du jeu avec la formule de l'énoncé, mais un argument simple et intuitif permet de conclure que l'entropie diminue: avec un pion de moins sur l'échiquier, le nombre de positions différentes qu'on peut représenter est moindre que celui qu'on peut représenter au départ. Si vous n'êtes pas convaincus, pensez maintenant à la situation d'une fin de partie où il n'y a plus que les deux rois sur l'échiquier et un pion, par exemple. Dans ce cas, l'entropie du jeu est plus petite qu'au départ, car le nombre de positions différentes qu'on peut former avec ces 3 éléments sur l'échiquier est clairement bien moindre qu'avec toutes les pièces.

Note: Si vous avez déjà entendu parler du second principe de la thermodynamique, qui dit que l'entropie d'un système fermé augmente au cours du temps, vous avez peut-être été surpris par le résultat ci-dessus. Remarquez cependant qu'un jeu d'échecs n'est pas un système fermé, car au fur et à mesure du jeu, on pose les pièces capturées sur le bord de l'échiquier.

2 Comparaisons d'entropies

a) On indique ci-dessous en **gras** le mot avec la plus grande entropie:

1. EPFL et EEPPFFLL: Les entropies sont égales (et valent 2) car dans les deux cas, les 4 lettres E, P, F et L apparaissent toutes avec probabilité 1/4.

2. AAAH et **HAHA**: L'entropie du premier mot vaut $(3/4)\log_2(4/3) + (1/4)\log_2(4) \simeq 0.81$, tandis celle du deuxième vaut 1. On peut aussi voir qu'il y a moins de possibilités de former des mots différents avec les 4 lettres AAAH qu'avec les 4 lettres HAHA, donc moins d'entropie dans le premier mot.

3. **MEDITERRANNEE** et MEDETERRENNEE: L'entropie du premier mot est clairement plus grande, car le nombre de lettres dans chaque mot est le même, mais toutes les voyelles sont remplacées par des E dans le second mot.

4. ABB et **ABBA**: L'entropie du premier mot vaut $(2/3)\log_2(3/2) + (1/3)\log_2(3) \simeq 0.91$, tandis que celle du second vaut 1.

5. **ACD** et ACDC: L'entropie du premier mot vaut $\log_2(3) \simeq 1.58$, tandis que celle du second vaut 1.5.

6. ABR et **ABRI**: L'entropie du premier mot vaut $\log_2(3)$, tandis que celle du second vaut $\log_2(4)$.

7. CALC et **CALCUL**: L'entropie du premier mot vaut 1.5, tandis que celle du second vaut $(2/3)\log_2(3) + (1/3)\log_2(6) \simeq 1.92$.

b) La réponse est non. Il se peut que l'entropie augmente (par exemple, avec ABB et ABBA) ou diminue (par exemple, avec ACD et ACDC).

c) Ici, la réponse est oui (par exemple, avec ABR et ABRI). La démonstration formelle de ce fait dépasse le cadre de ce cours, mais une intuition simple est la suivante: rajouter une lettre qui ne fait pas partie de la séquence amène quelque chose de *nouveau* à celle-ci et augmente donc le nombre de possibilités de former des mots différents, ce qui implique une augmentation de l'entropie.

3 Un peu de magie noire

a) Calcul de l'entropie de la séquence (que l'on va appeler $\mathcal{X}$ comme dans le cours): le A a une probabilité d'apparition de $\frac{5}{12}$, le V et le D de $\frac{2}{12}$ et les 3 lettres restantes de $\frac{1}{12}$. Donc

$$\begin{aligned} H(\mathcal{X}) &= \frac{5}{12}\log_2\left(\frac{12}{5}\right) + 2\,\frac{2}{12}\log_2\left(\frac{12}{2}\right) + 3\,\frac{1}{12}\log_2(12) \\ &= \log_2(12) - \frac{5}{12}\log_2(5) - \frac{1}{3}\log_2(2) = \frac{5}{3} + \log_2(3) - \frac{5}{12}\log_2(5) \end{aligned}$$

ce qui donne numériquement: $H(\mathcal{X}) \simeq 2.28$.

b) Avec l'algorithme de Shannon-Fano, on peut trouver les dictionnaires suivants:

lettre	nb app.	1) nb Q	mot de code	2) nb Q	mot de code	3) nb Q	mot de code
A	5	2	11	2	11	1	1
V	2	2	10	2	10	3	011
D	2	3	011	2	01	3	010
K	1	3	010	3	001	3	001
E	1	3	001	4	0001	4	0001
R	1	3	000	4	0000	4	0000

Avec les dictionnaires 1) et 2), la séquence codée contient 29 bits; avec le dictionnaire 3), la séquence codée contient 28 bits.

c) Avec l'algorithme de Huffman, on trouve le code suivant:

lettre	nb apparitions	mot de code
A	5	1
V	2	011
D	2	010
K	1	001
E	1	0001
R	1	0000

et la séquence codée contient 28 bits. On peut aussi trouver des dictionnaires différents suivant l'arbre que l'on construit, mais la longueur de la séquence codée reste invariablement de 28 bits dans ce cas.

d) Avec l'algorithme de Shannon-Fano, la longueur moyenne du code $L(\mathcal{C}_{SF}) \simeq 2.42$ ou 2.33. Avec l'algorithme de Huffman, la longueur moyenne du code $L(\mathcal{C}_H) \simeq 2.33$, et l'entropie de la séquence vaut $H \simeq 2.28$. On vérifie donc bien les inégalités du cours:

$$H(\mathcal{X}) \leq L(\mathcal{C}_H) \leq L(\mathcal{C}_{SF}) \leq H(\mathcal{X}) + 1$$

4 A la recherche d'un trésor

a) On doit chaque fois se mettre au point sur la grille qui est au milieu des possibilités qui nous restent, de sorte à diviser cet ensemble de possibilités en quatre parties égales; vu qu'il y a 64 possibilités au départ, le nombre de questions à poser est $\log_4(64) = 3$.

b) (5,5), (7,7), (6,8) (et la suite des réponses correspondantes de l'oracle est NE, NO, SE).

c) Deux possibilités:

- En vous basant sur la suite des questions ci-dessus, vous encodez chaque réponse obtenue avec 2 bits, p.ex.: NE=00, NO=01, SE=10, SO=11 (on peut aussi dire, N=0, S=1, O=0, E=1 d'ailleurs, mais se rappeler alors qu'on indique d'abord la direction nord-sud avant la direction est-ouest). Dans le cas présent, la séquence est donc 000110.

- Sans rapport avec le jeu des questions ci-dessus, on peut aussi simplement encoder la position de la case avec 6 bits également: 3 bits pour la position horizontale (de 1 à 8: plus précisément, on va encoder le nombre moins 1, comme dans l'exercice sur le codage par plages: ainsi, 000 encode 1, 001 encode 2, jusqu'à 111 qui encode 8), et 3 bits pour la position verticale. Avec cette représentation, la position du trésor ci-dessus est (6,7), et donc l'encodage est 101110 (on utilise la convention que l'abscisse vient avant l'ordonnée: pas besoin donc d'utiliser un bit pour ça: ça fait partie de la manière dont on définit notre dictionnaire).