EPFL - Information, Calcul et Communication - Partie théorique

# Série 11

## 1 Comparaisons d'entropies

Pour chaque paire de mots ci-dessous, estimer lequel des deux mots a la plus grande entropie, ou s'ils ont la même entropie. Pour cela, on peut bien sûr à chaque fois calculer les entropies des deux mots pour répondre, mais on peut aussi essayer de répondre sans faire de calculs, en raisonnant sur les probabilités d'apparition des lettres dans chaque mot (autre indice: à nombre égal de lettres, le mot qui donne le plus de possibilités de créer d'autres mots en réutilisant ses lettres est celui qui a la plus grande entropie).

1. EPFL et EEPPFFLL
2. MEDITERRANNEE et MEDETERRENNEE
3. AAAH et HAHA
4. ABB et ABBA
5. ACD et ACDC
6. ABR et ABRI
7. CALC et CALCUL

## 2 Est-ce que l'entropie augmente ou diminue?

**a)** Considérons une séquence de lettres le longueur $N$ finie. Est-il *toujours* vrai que si on ajoute à cette séquence une lettre qui fait déjà partie de la séquence, alors l'entropie de la nouvelle séquence (de longueur $N+1$) diminue?

**b)** Considérons une séquence de lettres le longueur $N$ finie. Est-il *toujours* vrai que si on ajoute à cette séquence une lettre qui ne fait pas partie de la séquence, alors l'entropie de la nouvelle séquence (de longueur $N+1$) augmente?

*Note:* Pour répondre à chacune de ces deux questions ci-dessus, vous pouvez vous aider de l'exercice précédent. A chaque fois, il vous faut soit vous convaincre que c'est toujours vrai, à l'aide de plusieurs exemples (car la démonstration formelle est difficile et dépasse le cadre de ce cours), soit trouver un contre-exemple à l'affirmation énoncée.

## 3 Un peu de magie noire

**a)** Calculez l'entropie de la séquence de lettres suivante (sans l'espace):

AVADA KEDAVRA

- en écrivant d'abord le résultat sous la forme $a + b\,\log_2(3) + c\,\log_2(5)$, où $a, b, c$ sont des fractions de nombres entiers;

- puis en calculant le résultat numérique ou en l'approximant avec $\log_2(3) \simeq 1.58$ et $\log_2(5) \simeq 2.32$.

**b)** Créez un dictionnaire pour cette même séquence de lettres à l'aide de l'algorithme de Shannon-Fano. Combien de bits utilisez-vous pour représenter la séquence? Essayez différentes versions de l'algorithme et comparez.

**c)** Créez ensuite un dictionnaire à l'aide de l'algorithme de Huffman. Combien de bits utilisez-vous pour représenter la séquence? A nouveau, essayez différentes versions de l'algorithme et comparez.

**d)** Comparez les résultats obtenus aux points a), b) et c). Ceci est-il cohérent avec ce que vous avez appris au cours?

# 4 De bien mauvais dicos

Pour encoder le mot “BANANA” sous forme de 0 et de 1, six minions ont proposé les dictionnaires suivants:

**a)**

| A | N | B |
|---|---|---|
| 1 | 11 | 111 |

**b)**

| A | N | B |
|---|---|---|
| 0 | 1 | 11 |

**c)**

| A | N | B |
|---|---|---|
| 0 | 10 | 110 |

**d)**

| A | N | B |
|---|---|---|
| 00 | 01 | 10 |

**e)**

| A | N | B |
|---|---|---|
| 00 | 01 | 1 |

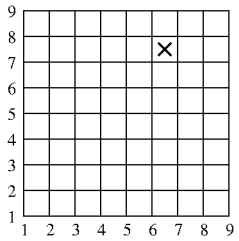
**f)**

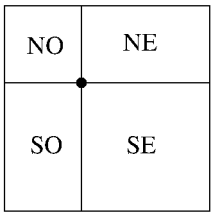
| A | N | B |
|---|---|---|
| 0 | 01 | 001 |

Mais Grou n’est satisfait d’aucun d’entre eux. Pouvez-vous expliquer le(s) défaut(s) de chacun? Et aussi proposer un meilleur dictionnaire?

# 5 A la recherche d’un trésor

Vous êtes à la recherche d’un trésor, dont la position est une case choisie au hasard et indiquée par une croix sur le plan à gauche ci-dessous. Pour trouver la position du trésor, vous devez choisir un point sur la grille et poser la question: dans quelle direction se trouve le trésor? Un oracle vous indiquera alors s’il est au nord-est (NE), au nord-ouest (NO), au sud-est (SE) ou au sud-ouest (SO), comme indiqué par exemple à droite ci-dessous. Bien sûr, il vous faudra interroger l’oracle en plusieurs points de la grille pour localiser le trésor.



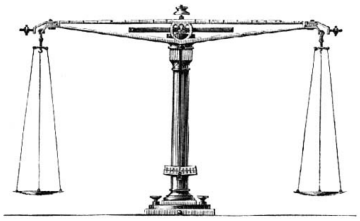

**a)** Si vous utilisez une stratégie optimale, combien de questions faudra-t-il pour localiser le trésor?

**b)** Dans l’exemple ci-dessus à gauche, quelle est la suite des points où vous interrogerez l’oracle avec votre stratégie optimale? (utilisez la notation $(a, b)$ pour un point, avec $a, b$ allant de 1 à 9, $a$ dénotant l’abscisse et $b$ l’ordonnée du point).

**c)** Proposez un système *efficace* d’encodage avec des bits pour transmettre l’information de l’emplacement du trésor à un ami. En se basant sur votre système, quelle sera la séquence de bits que vous transmettrez à votre ami dans l’exemple ci-dessus à gauche?

## 6 Pour le plaisir: minimiser le nombre de pesées*

**a)** Vous avez devant vous 9 pièces de monnaie, identiques en apparence, mais quelqu'un vous dit que l'une d'elles est fausse et pèse un peu moins lourd que les autres. Pour identifier la fausse pièce, vous disposez d'une balance:



avec laquelle il est possible d'effectuer des pesées. Le résultat de chaque pesée peut être "la balance penche à gauche", "la balance penche à droite" ou "la balance reste stable".

En admettant que vous utilisez une stratégie optimale, quel est le nombre *minimum* de pesées qu'il vous faut exécuter pour identifier la fausse pièce? Et quelles sont ces pesées?

*Indication:* Commencez par le problème plus simple où vous n'avez que 3 pièces devant vous.

**b)** Supposons maintenant que vous ayez 4 pièces devant vous, et que l'on vous indique qu'*au plus* une d'entre elles est fausse, *sans vous dire si celle-ci est plus légère ou au contraire plus lourde que les autres.* Pour vous aider, vous avez cette fois une pièce additionnelle dans votre poche que vous savez être vraie et que vous pouvez utiliser pour les pesées.

En admettant que vous utilisez une stratégie optimale, quel est le nombre minimum de pesées qu'il vous faut exécuter (et à nouveau, quelles pesées effectuerez-vous?) pour être en mesure de répondre aux trois questions suivantes (prises ensemble): y a-t-il une fausse pièce? le cas échéant, quelle est-elle? et est-elle plus lourde ou plus légère que les autres?