EISTI **Année scolaire 2007-2008**

DÉPARTEMENT "SYSTÈMES INFORMATIQUES FORMELS ET INTELLIGENTS"

THÉORIE DE L'INFORMATION

Devoir surveillé du 12 juin 2008. *Durée 1h30.*

Document autorisé : *polycopié du cours imprimé uniquement. Les* **calculatrices** *sont admises.*

Exercice 1. La loterie à Babylone

1. "Désormais, les acheteurs de rectangles numérotés avaient la chance de gagner une certaine somme ou de payer une amende parfois considérable. Ce léger danger (il y avait un numéro funeste tous les trente numéros favorables) éveilla naturellement l'intérêt du public. Les Babyloniens se livrèrent au jeu."

   Jorge Luis Borges, La loterie à Babylone

   (a) Quelle est la quantité d'information associée au gain lors d'un tirage de la loterie babylonienne ?

   (b) Quelle est la quantité d'information associée à l'amende ?

   (c) Quelle est l'entropie du jeu ?

   (d) Quelle est la valeur maximale d'entropie d'un jeu binaire ?

2. "Il y a aussi des tirages impersonnels, d'une intention indéfinie ; celui-ci ordonnera de jeter un saphir de Taprobane dans les eaux de l'Euphrate ; cet autre, de lâcher un oiseau du haut d'une tour ; cet autre de retirer tous les siècles un grain de sable à la plage, ou de l'y ajouter."

   Jorge Luis Borges, La loterie à Babylone

   En supposant que les probabilités des quatres sentences sont les suivantes :

saphir	oiseau	retirer un grain de sable	ajouter un grain
0.4	0.4	0.1	0.1

   calculer l'entropie d'une telle loterie.

Exercice 2. Soit une source d'alphabet $\Omega = \{a, b, c, d, e, f\}$ et de distribution de probabilité suivante :

X	a	b	c	d	e	f
P_X	0.25	0.15	0.1	0.35	0.1	0.05

1. Calculer l'entropie de cette source.
2. Quelle est la borne inférieure pour la longueur moyenne d'un code déchiffrable pour cette source ? Existe-t-il un code absolument optimal pour cette source ? Justifier votre réponse.
3. Construire un code de Huffman pour cette source.
4. Calculer la longueur moyenne du code obtenu.

Exercice 3. Soit une source binaire X d'alphabet $\Omega_X = \{0, 1\}$. On suppose que la distribution de probabilité de X est connue :

$$p_X(0) = p, \quad p_X(1) = 1 - p$$

Les symboles émis sont envoyés via un canal de transmission ayant des perturbations aléatoires. On associe à l'expérience d'observation du symbole reçu la variable aléatoire Y qui peut prendre trois valeurs : $\Omega_Y = \{-1, 0, 1\}$. La valeur -1 correspond au cas où le système n'est pas capable d'identifier un 0 ou un 1 à la sortie. On suppose que la matrice de transition du canal est la suivante :

$P(Y|X) =$

$X \backslash Y$	0	-1	1
0	0.8	0.2	0
1	0	0.2	0.8

1. En utilisant la matrice de transition $P(Y|X)$ et la distribution de probabilité de X, P_X, calculer la distribution de probabilité conjointe $P(X, Y)$. En déduire P_Y, la distribution marginale du récepteur, Y et la distribution conditionnelle $P(X|Y)$.
2. Exprimer l'entropie $H_X(p)$ comme fonction du paramètre p.
3. Calculer $H(X|Y)$ en fonction du paramètre p.
4. Calculer $I(X; Y) = H_X(p) - H(X|Y)$. Montrer que $I(X; Y) = 0.8 H_X(p)$.
5. En déduire la valeur de la capacité du canal définie par

$$C = \max_{P_X} I(X; Y)$$

où le maximum est pris selon toutes les distributions possibles de la source X.

Exercice 4. Entropie et les algorithmes de tri

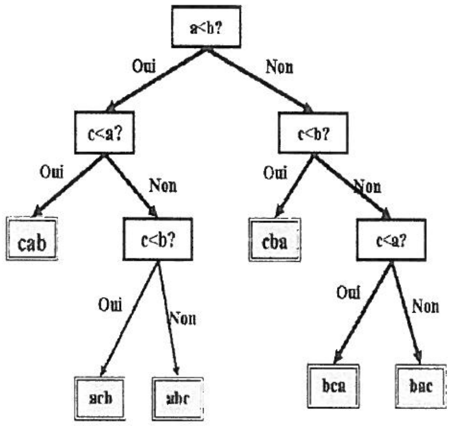
L'objectif de l'exercice est d'établir la borne inférieure pour la complexité des algorithmes de tri à base de comparaisons.

Soit un ensemble de n nombres deux à deux distincts. On cherche à ordonner l'ensemble dans l'ordre croissant à l'aide de comparaisons. On cherche donc une permutation particulière des n nombres. Toute comparaison est une question binaire et on peut représenter une succession de comparaisons sous forme d'arbre dont les feuilles sont les permutations possibles des n nombres.

1. Soient $n = 3$ et $A = \{a, b, c\}$ l'ensemble à trier.

   (a) Quel est le nombre de permutations différentes possibles ? En supposant que toutes les permutations sont équiprobables, quelle est l'entropie de l'ensemble de permutations ?

   (b) Voici (voir la figure ci-dessous) un exemple d'algorithme de tri pour un ensemble de trois éléments. Calculer le nombre moyen de comparaisons.



(c) Quelle est la borne inférieure pour le nombre moyen de comparaisons ? **Indication.** On peut assimiler l'arbre de l'algorithme à un arbre de code binaire et le nombre moyen de questions à la longueur moyenne de mots de code.

2. Considérons le cas général d'un ensemble de n éléments à trier.
   (a) En supposant que toutes les permutations sont toujours équiprobables, quelle est l'entropie du problème ?
   (b) Quelle est alors la borne inférieure pour le nombre moyen de comparaisons de tout algorithme de tri ?
   (c) En utilisant l'approximation de Stirling

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \quad n \to \infty$$

   Montrer que la complexité en moyenne C_m et la complexité au pire, C_p de tout algorithme de tri à base de comparaisons vérifient :

$$C_p \geq C_m \geq n \log n - n \log e$$

Pour finir en beauté.

"Comme tous les hommes de Babylone, j'ai été proconsul ; comme eux tous, esclave ; j'ai connu comme eux tous l'omnipotence, l'opprobre, les prisons. Regardez : à ma main droite il manque l'index. Regardez : cette déchirure de mon manteau laisse voir sur mon estomac un tatouage vermeil : c'est le deuxième symbole, Beth. Les nuits de pleine lune, cette lettre me donne le pouvoir sur les hommes dont la marque est Ghimel, mais elle me subordonne à ceux d'Aleph, qui les nuits sans lune doivent obéissance à ceux de Ghimel. Au crépuscule de l'aube, j'ai égorgé des taureaux sacrés devant une pierre noire. Toute une année de lune durant, j'ai été déclaré invisible : je criais et on ne me répondait pas, je volais le pain et je n'étais pas décapité. J'ai connu ce qu'ignorent les Grecs : **l'incertitude**."

Jorge Luis Borges, La loterie à Babylone

BONNES VACANCES !