Devoir Maison 1

Les notations qui ne sont pas rappelées sont celles du cours. Vous êtes autorisés à réfléchir à plusieurs sur les exercies (pas plus de deux ou trois). Cependant, la rédaction doit être individuelle et chacun doit rendre une copie. Précisez les noms de vos collaborateurs.

Exercice 1

On considère une source ayant un alphabet de $M = 4$ symboles ayant des probabilités $p_1 \geq p_2 \geq p_3 \geq p_4 > 0$.

1. Quels sont les arbres de Huffman possibles si $p_1 = p_3 + p_4$?

2. Quelle est la plus grande valeur de p_1 telle que $p_1 = p_3 + p_4$? La plus petite ?

On note respectivement $p_{\max}$ et $p_{\min}$ ces valeurs.

3. Montrer que si $p_1 > p_{\max}$, alors tout code de Huffman a un mot-code de longueur 1.

4. Montrer que si $p_1 < p_{\max}$, alors tous les mots-code ont longueur 2 pour tout code de Huffman.

5. On suppose $M > 4$. Trouver la plus petite valeur p'_{max} telle que $p_1 > p'_{\max}$ garantisse qu'un code de Huffman aura un mot-code de longueur 1.

Exercice 2 Codage de Huffman pour des lettres équiprobables

Une variable aléatoire prend ses valeurs dans un alphabet de K lettres et chaque lettre a la même probabilité. Ces lettres sont encodées dans des mots binaires de façon à minimiser la longueur moyenne des mots-code. On définit l'entier j et le réel $1 \leq x < 2$ tel que $K = x2^j$.

1. Montrer que tous les mots-code sont de longueur j ou $j + 1$.

2. Quelle est la longueur moyenne d'un mot-code ?

Exercice 3 Le paradoxe de St Petersbourg

On considère le jeu suivant : le joueur paye un droit d'entrée de c euros, et reçoit 2^k euros avec probabilité 2^{-k}.

1. Quel est le gain moyen du joueur ?

On pourrait donc penser que quelque soit c, c est un juste prix... On suppose maintenant que le joueur peut jouer une fraction de c : s'il paye $c/2$, alors il reçoit la récompense $X/2$ où $\mathbf{P}(X = 2^k) = 2^{-k}$. Supposons que $X_1, X_2, \ldots,$ sont i.i.d. distribuées comme X et sont les gains successifs lorsque c est payé. On suppose que le joueur réinvestit toute sa fortune à chaque fois. Au début sa fortune est 1. Après un tout, sa fortune est donc X_1/c.

2. Donne une expression de la fortune S_n du jour après n tours. Montrer que S_n tend vers 0 ou $+\infty$ selon que $c < c^*$ ou $c > c^*$. Quelle est la valeur de c^* ?

On appelle c^* le juste prix d'entrée.

On admet que le joueur peut garder une partie de sa fortune. On suppose qu'à chaque tour il garde une proportion $\bar{b} = 1 - b$ de sa fortune.

3. Donner une nouvelle expression pour la fortune du joueur après n tours. On note toujours S_n cette quantité.

4. À partir de quelle valeur de c le joueur choisit-il de miser toute sa fortune ?

5. Si le gain est maintenant tel que $\mathbf{P}(X = 2^{2^k}) = 2^{-k}$, faut-il miser toute sa fortune ?

`Exercice 4`

Soit X une variable aléatoire à valeurs dans $\mathbb{N} = \{0, 1, 2, \ldots\}$ et de distribution $(p(x))_{x\in\mathbb{N}}$. Un encodage binaire de X est une injection $\varphi : \mathbb{N} \to \{0,1\}^*$ de $\mathbb{N}$ dans l'ensemble des mots binaires finis (y compris le mot vide). La longueur moyenne de l'encodage φ est : $\ell(\varphi) = \sum_{x\in\mathbb{N}} p(x)|\varphi(x)|$, où $|\varphi(x)|$ est la longueur du mot $\varphi(x)$.

Dans certain cas, s'il existe un symbole encodant la fin d'un message, il n'est pas nécessaire de considérer des codes ayant la propriété du préfixe. On définit donc

$$\mathcal{L}_{1:1}(X) = \min\{\ell(\varphi),\ \varphi \text{ est un encodage de } X\}.$$

1. Montrer que $\mathcal{L}_{1:1}(X) \leq H(X)$. Donner un encodage optimal.

2. Montrer que pour toute variable aléatoire U à valeurs dans $\mathbb{N}$, on a $H(U) \leq \log(\mathbf{E}[U] + 1) + \log e$.

3. En déduire que $H(X) \leq \mathcal{L}_{1:1}(X) + \log(\mathcal{L}_{1:1}(X)+1) + \log e$, puis donner une borne inférieure pour $\mathcal{L}_{1:1}(X)$. On pourra utiliser l'axiome (A8) du TD1 dans la définition axiomatique de l'entropie, avec une partition bien choisie.

On considère maintenant le problème suivant : (X, Y) est un couple de variables aléatoires à valeurs dans l'espace dénombrable $\mathcal{X} \times \mathcal{Y}$ de distribution $p(x, y)$. Alice connaît X, Bob connaît Y et veut connaître X. On suppose qu'Alice peut communiquer vers Bob sans erreur ; que Bob doit pouvoir déterminer la fin d'un message d'Alice ; qu'Alice et Bob se sont mis d'accord sur un protocole déterministe qui peut dépendre de p. Le but est de trouver le nombre moyen de bits qu'Alice doit envoyer à Bob. On définit le support $S = \{(x, y), p(x, y) > 0\}$ et $x \neq x'$ sont ambigus s'il existe y tel que $(x, y), (x', y) \in S$. Un protocole pour des entrées restreintes est une fonction $\varphi : \mathcal{X} \to \{0,1\}^*$ telle que pour x et x' ambigus, $\varphi(x)$ n'est pas un préfixe (au sens large) de $\varphi(x')$. On définit le nombre de bits moyens pour φ comme précédement : $\ell(\varphi) = \sum_{x\in\mathbb{N}} |\varphi(x)|p(x)$.

On définit alors $\bar{L} = \min\{\ell(\varphi),\ \varphi \text{ est un protocole pour entrées restreintes } (X, Y)\}$.

4. Montrer que $H(X|Y) \leq \bar{L} \leq H(X) + 1$, que ces bornes sont les meilleures possibles et qu'elles peuvent être arbitrairement éloignées l'une de l'autre.

Clairement, $\bar{L}$ ne dépend de (X, Y) que par S et la distribution $p(x)$ (les valeurs de $p(y|x)$ n'interviennent pas dans les définitions). On définit donc le graphe G dont l'ensemble des sommets est $\mathcal{X}$ et deux sommets distincts x et x' sont connectés si ils sont ambigus. Le graphe probabiliste (G, X) est défini par le graphe G et la distribution de probabilité sur ses sommets $p(x)$. $\bar{L}$ ne dépend que de (G, X).

Si X est une variable aléatoire à valeur dans $\mathcal{X}$ et c est une fonction définie sur X alors $c(X)$ est une variable aléatoire d'entropie :

$$H[c(X)] = -\sum_{\gamma\in c(X)} p[c^{-1}(\gamma)] \log p[c^{-1}(\gamma)],$$

où c^{-1} est l'inverse de c et la probabilité d'un ensemble est la somme des probabilités de ses éléments. On définit l'entropie chromatique d'un graphe probabiliste (G, X) par

$$H(G, X) = \min\{H[c(X)], c \text{ est un coloriage de } G\}.$$

5. Donner l'entropie chromatique pour : le graphe vide, le graphe complet, le pentagone avec la distribution uniforme sur ses sommets, le cycle avec $p_0 = 0.3$, $p_1 = p_2 = p_3 = 0.2$ et $p_4 = 0.1$.

6. Un protocole pour des entrées non-restreintes est une fonction $\varphi : \mathcal{X} \to \{0,1\}^*$ telle que pour $x \neq x'$, $\varphi(x)$ n'est pas un préfixe propre de $\varphi(x')$ et si x et x sont ambigus, $\varphi(x) \neq \varphi(x')$. Soit $\bar{\mathcal{L}} = \min\{\ell(\varphi) : \varphi \text{ est un protocole pour entrées non-restreintes } (X,Y)\}$. Montrer que $H(G,X) \leq \bar{\mathcal{L}} \leq H(G,X) + 1$.

7. En utilisant la première partie de l'exercice, montrer que

$$H(G,X) - \log[H(G,X)+1] - \log e \leq \bar{L} \leq H(G,X) + 1.$$