TD de TIC

Mohamed Kadhem Karray

26 mars 2024

Table des matières

1 TD3

1.1 Exercice 1

Soit c un code PF. Soit $x_1, \ldots, x_k, y_1, \ldots, y_l$ pour lesquels il y a égalité. Alors, soit $i = \min\{i \in \mathbb{N} \mid x_i \neq y_i\}$. Alors en ne comptant que les i premiers caractères du mot de désection, selon la longueur des codes, ou bien $c(x_1, \ldots x_{i-1})$ est un préfixe de $c(y_1, \ldots, y_{i-1})$ ou bien l'inverse.

1.2 Exercice 2

1.2.1 Question 1

c est un code PF.

1.2.2 Question 2

Pour déchiffrer le code, il suffit de regarder le premier 0 que l'on trouve. On choisit entre $1, 2$ et 3 pour le déchiffrer selon la valeur modulo 3 du nombre de 1 qui le suivent.

1.3 Exercice 3

J'ai une tête de cloche à fromage ouuuuuuuuuuuu ?

1.4 Exercice 4

1.4.1 Question 1

On considère $\mathcal{U}_m = \{u \in \mathcal{U} \mid l(u) \leq m\}$. On a alors, puisque $\mathcal{U}_m$ est fini, l'existence de $l_{max,m} = \max\{l(u) \mid u \in \mathcal{U}_m\} \leq m$.
On considère ensuite c_m le code c tronqué à $\mathcal{U}_m$. On a : La suite des $\mathcal{U}_m$ est croissante et on a $\limsup \mathcal{U}_m = \mathcal{U}$. Donc, par passage à la limite supérieure, on a bien

$$\sum_{u \in \mathcal{U}} D^{-l(u)} \leq 1$$

1.4.2 Question 2

On retire, à sectionir de la profondeur $l(1)$ une proportion $\alpha_1 D^{-l(1)}$ des branches. On définit une mesure μ sur les fils de l'arbre comme la proportion des feuilles de l'arbre qu'ils recouvrent.
On note A_i l'union des fils enracinés à profondeur $l(i)$ qu'on a utilisés pour du codage de longueur $l(i)$. On a $\mu(A_i) = \alpha_i D^{-l(i)}$. Notre codage complet est l'union des A_i. On a alors :

$$\mu\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) = \sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{n} \alpha(i) D^{-l(i)} = \sum_{u \in \mathcal{U}} D^{-l(u)} \leq 1$$

Donc on n'a pas de recouvrement des arbres.
za

1.5 Exercice 5

1.5.1 Question 1

On code chaque nombre en base 2 sur k bits. On demande pour chaque bit si le nombre a la même valeur. Bah on a trouvé.

1.5.2 Question 2

On utilise le code $c_2(k)$ sur les K objets. Puisque le code est PF, on peut juste regarder pour tout i le i-ème symbole, et en fonction de ce qu'on a lu sur les $i-1$ premiers, lorsqu'on arrive à la fin du code, on a trouvé le résultat. On a $E[l(c_2(U))]$ questions en moyenne et donc on a bien :

$$H_2(U) \leq E[l(c_2)(U)] \leq H_2(U) + 1$$

Pour tout $u \in \{1, \ldots, K\}$, on utilise $\lceil \log_2(p_u) \rceil$ questions.

1.5.3 Question 3

Le nombre moyen de questions d'un questionnaire est équivalent à la longueur moyenne du codage UD qu'il représente. Par inégalité de Kraft et inégalité de Gibbs, on a :

$$\begin{aligned}
E[l(c)(k)] =& \sum_{k=1}^{K} p_k l(c(k)) \\
=& -\sum_{k=1}^{K} p_k \log_2(2^{-l(c(k))}) \\
\geq& -\sum p_k \log_2 q_k \\
\geq& -\sum p_k \log_2 p_k \\
=& H_2(U)
\end{aligned}$$

Donc le questionnaire a en moyenne au moins $H_2(U)$ questions.

1.5.4 Question 4

On considère $U_1, \ldots, U_m$ des réalisations i.i.d. de U.

1. Avec c_2, en moyenne, on va utiliser :

$$H_2(U) \leq \frac{1}{m} E[l(c_2(U_1)) + \ldots + l(c_2(U_m))] \leq \frac{H_2(U) + 1}{m}$$

2. Avec un code arbitraire, on va utiliser :

$$H_2(U) \leq \frac{1}{m} E[l(c(U_1)) + \ldots l(c(U_m))]$$

1.5.5 Question 5

On passe à la limite dans la question précédente.

2 TD4

2.1 Exercice 1

On a :

$$\begin{aligned}
E[X] &= \int_0^\infty x f(x) \, \mathrm{d}x \\
&= \int_0^\infty [x f_X(t)]_0^t \, \mathrm{d}t \\
&= \int_0^\infty \int_0^t f_X(t) \, \mathrm{d}x \, \mathrm{d}t \\
&= \int_0^\infty \int_x^\infty f_X(t) \, \mathrm{d}t \, \mathrm{d}x && \text{Fubini Tonelli} \\
&= \int_0^\infty P(X \geq x) \, \mathrm{d}x
\end{aligned}$$

2.2 Exercice 2

On a :

$$\begin{aligned} P(Y = y, Z = z|X = x) =& P(Y = y|Z = z, X = x) \times P(Z = z|X = x) \\ =& f(x, y) \times P(Z = z|X = x) \end{aligned}$$

Donc on a bien l'indépendance car

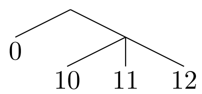
$$P(Y = y|X = x) = \sum_{z \in Z(\Omega)} P(Y = y|X = x, Z = z) = f(x, y)$$

2.3 Exercice 3

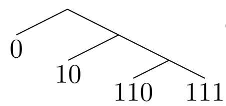
1. On sépare à chaque étape les pièces en 3. Puisqu'une unique pièce est plus lourde, une balance suffit à trouver dans quel groupe la pièce la plus lourde est : Si l'un des ensembles pesés est plus lourd, c'est dans celui là, sinon, s'il y a égalité c'est dans le troisième.
2. Chaque pesée correspond à une manière de séparer les pièces en 3. En sectioniculier, l'arbre des possibilités correspondant à une stratégie arbitraire va être de hauteur au moins $\log_3(n)$. La hauteur de l'arbre correspondant bien au nombre moyen de pesées...
3. On utilise la stratégie de la première question. On n'utilise alors jamais plus de $\lceil \log_3(n) \rceil$ pesée. Si $n = 3^k$, on revient au premier cas avec $\lfloor \log_3(n) \rfloor = k$ pesées.
4. On calcule $H(U|Y) = \log_3(n) + \frac{\varphi(p)}{\log(3)}$ où $\varphi(p) = 2p \log p + (1 - 2p) \log(1 - p)$
5. On peut tester en deux pesées lors de la première étape si la pièce différente est plus lourde ou plus légère que les autres. En effet, si on sépare en trois ensembles A, B, C. Ou bien : $A < B$ et $B = C$ et la pièce est plus légère et dans A, ou bien $A < B$ et $B > C$ et la pièce est plus lourde et dans B. À renommage près, on applique la stratégie de 1) et on a le résultat. L'entropie dit qu'on nécessiterait $k + \log_3 2$ pesées en moyenne, ce qui n'est pas entier. On a donc une solution optimale.

2.4 Exercice 4

1. Il n'y a pas de feuilles manquantes à profondeur strictement inférieur à $x = max\,|c(u)|$ puisqu'on minimise la longueur moyenne des codes. On a par ailleurs $B \leq D - 2$ et $n + B = 1 + (D - 1)\alpha$.
2. Puisque tous les codes PF sont UD, le code de Huffman D-aire convient.
3. L'arbre ternaire qu'on obtient est :



On a alors $\bar{R_s} = 1.75$ et donc un taux de compression de $\frac{7/4}{\log_3 4} = 1.39$ Pour $D = 2$ on a l'arbre



ce qui donne $\bar{R_s} = 9/4$ et donc un taux de compression de 1.125.

4. Pour la distribution $p = (0.3, 0.2, 0.2, 0.15, 0.08, 0.07)$, on trouve

Binaire $L_C = 2.45$ et donc un taux de compression de 1.225

Ternaire $L_C = 1.65$ et donc un taux de compression de 1.012 environ

3 TD5

3.1 Exercice 1

On conditionne par rapport au système complet d'évènements $(T = k)$ pour $k \geq 1$. On a :

$$\begin{aligned}
\sum_{n=1}^{T} Y_n &= \sum_{n=1}^{+\infty} Y_n \mathbb{1}_{T \geq n} \\
\mathbb{E}\left[\sum_{n=1}^{T} Y_n\right] &= \mathbb{E}\left[\sum_{n=1}^{+\infty} Y_n \mathbb{1}_{T \geq n}\right] \\
&= \sum_{n=1}^{+\infty} \mathbb{E}\left[Y_n \mathbb{1}_{T \geq n}\right] \\
&= \sum_{n=1}^{+\infty} \mathbb{E}[Y_1] P(T \geq n) \\
&= \mathbb{E}[Y_1] \sum_{n=1}^{+\infty} P(T \geq n) \\
&= \mathbb{E}[Y_1]\mathbb{E}[T]
\end{aligned}$$

3.2 Exercice 2

En remarquant que si $U_1, \ldots, U_k$ forment le plus petit préfixe de U trouvé dans $\mathcal{C}$ (c'est le seul car le dictionnaire est valide), on peut écrire $U = U_1, \ldots, U_k, \tilde{U}$ où $\tilde{U}$ est une séquence infinie, on obtient immédiatement l'indépendance des W_i de l'indépendance des U_k. De même, du fait que les U_i suivent la même loi et du fait que le code est valide, on a bien l'identique distribution des W_i.

3.3 Exercice 3

On calcule la hessienne de f :

$$\begin{pmatrix} \frac{1}{x} & -\frac{1}{y} \\ -\frac{1}{y} & \frac{x}{y^2} \end{pmatrix}$$

Son polynôme caractéristique est :

$$\lambda^2 - \frac{\lambda}{x} - \frac{\lambda x}{y^2}$$

Ses valeurs propres (réelles) sont donc

$$\lambda_1 = 0 \text{ et } \lambda_2 = \frac{x^2 + y^2}{xy^2}$$

Ces deux valeurs propres sont positives donc f est convexe.

3.4 Exercice 4

On considère l'arbre (semi-infini) enraciné D-aire dont les noeuds sont étiquetés par la probabilité que la source émette la séquence de lettre qui amène de la racine à cette branche.
On observe que :

1. Tous les sommets dans le sous arbre fils de n'importe quel noeud sous moins probable que ce noeud
2. Tous les sommets dans l'arbre D-aire d'un dictionnaire valide pour la source sont aussi des noeuds avec la même probabilité dans notre arbre

Par définition d'un dictionnaire de Tunstall, un dictionnaire valide avec $M = D + \alpha(D-1)$ mots est un dictionnaire de Tunstall si et seulement ses $\alpha+1$ noeuds sont les $\alpha+1$ plus probables noeuds dans cet arbre. Ainsi, la somme des probabilités des noeuds est la même pour tous les dictionnaires de Tunstall à M mots. Mais, on sait que la profondeur moyenne des feuilles est égale à la somme des probabilités des noeuds (la probabilité de chaque noeud étant la somme des probabilités dans le sous-arbre).

4 TD6

4.1 Exercice 1

1. Chaque distribution est indexée de manière unique par n compteurs entre 0 et m, donc $|P_m| \leq (m+1)^n$.
2. En regroupant les indexations qui forment la même distribution, c'est à dire en précisant le dénombrement précédent, on obtient

$$|P_m| = \binom{m+n-1}{n-1}$$

3. Le nombre de suites de longueur n avec k_a fois la lettre a pour chaque $a \in \mathcal{U}$ est donné par le nombre de manière distinctes de permuter un multiensemble à n éléments :

$$|T_n(p)| = \binom{n}{k_{a_1} \cdots k_{a_n}} \frac{n!}{\prod_{a\in\mathcal{U}} k_a!}$$

4. Pour la borne supérieure, on a, en utilisant la co-entropie :

$$1 \geq \sum_{u^m \in T_m(p)} e^{-mH(p_{u^m}|p} = |T_m(p)|\, e^{-mH(p)}$$

Pour la borne inférieure, on a

$$\frac{1}{M} \log |T_m(p)| = H(q) + o(1)$$

Donc $T_m(p) \geq e^{m(H(q)+o(1))} \geq (m+1)^{-n} e^{mH(q)}$

4.2 Exercice 3

1. C'est l'inégalité de Gibbs :

$$D(p,q) = -\sum p_i \log q_i - \left(-\sum p_i \log p_i\right) = H(q,p) - H(p) \geq 0$$

Il y a égalité si et seulement si il y a égalité dans Gibbs, i.e. $p = q$ p.s.

2. Trivial de par la formule ci dessus.
3. Trivial de par la formule ci dessus.
4. Trivial par continuité de $H(q,\cdot)$ et de H.
5.

6. On a :

$$\begin{aligned}
I(X,Y) =& H(Y) - H(Y \mid X)\\
=& -\sum p_X(x) H(Y \mid X=x) - \sum p_Y(y)\log p_Y(y)\\
=& \sum p_X(x)\left(\sum p_{Y|X=x}(y)\log p_{Y|X=x}(y)\right) - \sum\left(\sum p_{(X,Y)}(x,y)\right)\log p_Y(y)\\
=& \sum p_X(x) p_{Y|X=x}(y)\log p_{Y|X=x}(y) - \sum p_{(X,Y)}(x,y)\log p_Y(y)\\
=& \sum p_{(X,Y)}(x,y)\log\frac{P_{(X,Y)}(x,y)}{p_X(x)} - \sum p_{(X,Y)}(x,y)\log p_Y(y)\\
=& \sum p_{(X,Y)}(x,y)\log\frac{p_{(X,Y)}(x,y)}{p_X(x)p_Y(y)}\\
=& D(P_{X,Y} \mid P_X \times P_Y)
\end{aligned}$$

<++>