

Théorie de l'information et du codage - Interrogation écrite

15 décembre 2000

Enoncés, solutions et explications

A. QCM - Théorie

Indiquer si les affirmations suivantes sont vraies ou fausses.

1. Notion d'indépendance d'événements

- (a) \emptyset est indépendant de tout autre événement

Vrai

Explication : pour tout événement A on a $P(\emptyset \cap A) = P(\emptyset) = 0 = P(\emptyset)P(A)$

- (b) " $A \perp B$ " \Rightarrow " $\neg A \perp \neg B$ ".

Vrai

Explication : D'une part on a $A \perp B \Rightarrow [P(A \cap B) = P(A)P(B) = P(A)(1 - P(\neg B))]$

D'autre part : $P(A \cap \neg B) + P(A \cap B) = P(A) \Rightarrow P(A \cap B) = P(A) - P(A \cap \neg B)$

On en déduit que $P(A \cap \neg B) = P(A)P(\neg B) \Rightarrow A \perp \neg B$, et donc $A \perp B \Rightarrow A \perp \neg B$.

Comme la relation \perp est symétrique, on en déduit aussi que $A \perp B \Rightarrow \neg A \perp B$.

La combinaison de ces deux propriétés donne la thèse.

- (c) $A \not\perp B$ et $B \not\perp C \Rightarrow A \not\perp C$

Faux

Contre-exemple : double pile ou face avec A face au premier lancer, C face au second lancer, B face aux deux lancers. On a $A \perp C$, et pourtant $A \not\perp B$ et $B \not\perp C$.

- (d) " $A \perp B$ " $\Leftrightarrow P(A \cap B) = P(A)P(B)$

Vrai

Par définition.

- (e) $A \perp B$ et $B \perp C \Rightarrow A \perp C$

Faux

Contre-exemple immédiat : double pile ou face, avec A l'événement "face au premier lancer", B l'événement "face au second lancer", C l'événement "pile au premier lancer"

2. Entropies conditionnelles

- (a) $H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}) < H(\mathcal{X}, \mathcal{Y})$

Faux

On peut avoir $H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}) = H(\mathcal{X}, \mathcal{Y})$.

- (b) $H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}) > H(\mathcal{X})$

Faux

Contre-exemple : si $\mathcal{X} = f(\mathcal{Z})$ et $\mathcal{Y} = g(\mathcal{X})$ on aura $H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}) = 0$, et comme $H(\mathcal{X}) \geq 0$ cela contredit la thèse.

- (c) $H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}) \leq H(\mathcal{X}, \mathcal{Y})$

Vrai

C'est une propriété de base (voir formulaire F.20)

$$(d) H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}, \mathcal{T}) \geq H(\mathcal{X}, \mathcal{Z} | \mathcal{T}) - H(\mathcal{Z} | \mathcal{T})$$

Vrai

En effet: On a $H(\mathcal{X}, \mathcal{Y}, \mathcal{Z} | \mathcal{T}) = H(\mathcal{Z} | \mathcal{T}) + H(\mathcal{X}, \mathcal{Y} | \mathcal{Z}, \mathcal{T})$ (règle de chaînage, valable aussi pour les entropies conditionnelles) et évidemment $H(\mathcal{X}, \mathcal{Y}, \mathcal{Z} | \mathcal{T}) \geq H(\mathcal{X}, \mathcal{Z} | \mathcal{T})$ (formule F.19, valable aussi pour les entropies conditionnelles).

3. Informations mutuelles

$$(a) I(\mathcal{X}; \mathcal{Y} | \mathcal{Z}) \geq I(\mathcal{X}; \mathcal{Y})$$

Faux

La règle générale, c'est qu'on ne peut rien dire en général en ce qui concerne l'effet du conditionnement sur l'information mutuelle de deux variables aléatoires.

Contre-exemple :

On dispose de deux urnes (une avec des billes bleues et rouges, l'autre avec des billes vertes et jaunes), et d'une pièce. On lance d'abord la pièce (variable \mathcal{Z}) pour choisir une des deux urnes, ensuite on tire une bille dans l'urne choisie (la variable \mathcal{X} désigne la couleur de cette bille), on la remet, et on choisit une deuxième bille (variable \mathcal{Y}) dans la même urne.

Ici, une fois \mathcal{Z} connue les variables \mathcal{X} et \mathcal{Y} sont indépendantes (et donc $I(\mathcal{X}; \mathcal{Y} | \mathcal{Z}) = 0$). Par contre, on a $I(\mathcal{X}; \mathcal{Y}) = 1$ Shannon (quels que soient les nombres relatifs de billes de chaque couleur dans chacune des deux urnes). En effet, remarquons tout d'abord que l'information apportée par \mathcal{X} sur \mathcal{Y} est la même que celle apportée par \mathcal{Z} sur \mathcal{Y} (la connaissance de l'urne dans laquelle on va tirer la seconde bille). Or, on a

$$H(\mathcal{Y}, \mathcal{Z}) = H(\mathcal{Y}) = H(\mathcal{Z}) + H(\mathcal{Y} | \mathcal{Z}).$$

Par conséquent on a aussi

$$I(\mathcal{X}; \mathcal{Y}) = I(\mathcal{Z}; \mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y} | \mathcal{Z}) = H(\mathcal{Z}) = 1$$

$$(b) I(\mathcal{X}; \mathcal{Y} | \mathcal{Z}) \leq I(\mathcal{X}; \mathcal{Y})$$

Faux

Contre-exemple : \mathcal{X} un texte binaire de longueur donnée, \mathcal{Y} un mot de passe, \mathcal{Z} le résultat de l'encryptage de \mathcal{X} au moyen de \mathcal{Y} avec une méthode DES ($\mathcal{Z} = DES(\mathcal{X}, \mathcal{Y})$). En supposant que le mot de passe et le texte sont choisis indépendamment on a $I(\mathcal{X}; \mathcal{Y}) = 0$, et pourtant, si on se donne \mathcal{Z} , la connaissance du mot de passe permet de retrouver le texte.

$$(c) I(\mathcal{X}; \mathcal{Y} | \mathcal{Z}) > I(\mathcal{X}; \mathcal{Y})$$

Faux

Puisque (a) est faux...

$$(d) I(\mathcal{X}; \mathcal{Y}, \mathcal{Z}) \geq I(\mathcal{X}; \mathcal{Y})$$

Vrai

Voir formule F.22

4. Indépendance de variables aléatoires (discrètes)

$$(a) \mathcal{X} \perp \mathcal{Y} \Rightarrow I(\mathcal{X}; \mathcal{Y}) = 0$$

Vrai

$\mathcal{X} \perp \mathcal{Y} \Rightarrow P(X_i, Y_j) = P(X_i)P(Y_j), \forall i, j \Rightarrow I(\mathcal{X}; \mathcal{Y}) = 0$ par application de F.6.

$$(b) \mathcal{X} \not\perp \mathcal{Y} \Rightarrow I(\mathcal{X}; \mathcal{Y}) > 0$$

Vrai

Par application de l'inégalité de Gibbs aux deux lois de probabilité $P(X_i, Y_j)$ et $P(X_i)P(Y_j)$, ou bien en utilisant le fait que la fonction H_n est une fonction strictement concave, ce qui a pour conséquence que $\mathcal{X} \not\perp \mathcal{Y} \Leftrightarrow H(\mathcal{X} | \mathcal{Y}) < H(\mathcal{X})$.

(c) $\mathcal{X} \perp \mathcal{Y} \Leftrightarrow H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y})$

Vrai

Comme conséquence de la propriété générale $H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - I(\mathcal{X}; \mathcal{Y})$ et des deux propriétés précédentes.

(d) $\mathcal{X} \not\perp \mathcal{Y} \Leftrightarrow H(\mathcal{X}|\mathcal{Y}) < H(\mathcal{X})$

Vrai

Cf. concavité stricte de la fonction entropie H_n .

5. Fonctions de variables aléatoires

(a) $\mathcal{Y} = g(\mathcal{X}) \Rightarrow H(\mathcal{Y}) \leq H(\mathcal{X})$

Vrai

Dans ce cas, $H(\mathcal{Y}|\mathcal{X}) = 0$; donc $H(\mathcal{Y}, \mathcal{X}) = H(\mathcal{X})$. Par ailleurs on a évidemment $H(\mathcal{Y}, \mathcal{X}) \geq H(\mathcal{Y})$.

(b) $\mathcal{Y} = g(\mathcal{X}) \Rightarrow \mathcal{Y} \perp \mathcal{Z}|\mathcal{X}, \forall \mathcal{Z}$

Vrai

En effet, $\mathcal{Y} = g(\mathcal{X}) \Rightarrow H(\mathcal{X}|\mathcal{Y}_j) = 0, \forall \mathcal{Y}_j$, et donc $I(\mathcal{X}; \mathcal{Z}|\mathcal{Y}_j) = 0, \forall \mathcal{Y}_j, \mathcal{Z}$ et donc $\mathcal{X} \perp \mathcal{Z}|\mathcal{Y}$.

(c) $\mathcal{Y} = g(\mathcal{X}) \Rightarrow H(\mathcal{Y}) \leq \log_2 |\mathcal{X}|$

Vrai

Puisque (a) est vrai et que $H(\mathcal{X}) \leq \log_2 |\mathcal{X}|$.

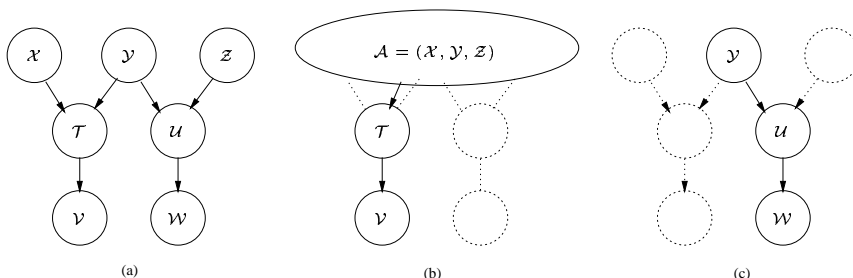
(d) $\mathcal{Y} = g(\mathcal{X}) \Rightarrow I(\mathcal{X}; \mathcal{Y}) = \min\{H(\mathcal{X}), H(\mathcal{Y})\}$

Vrai

En fait, puisque $\mathcal{Y} = g(\mathcal{X})$ on a $\min\{H(\mathcal{X}), H(\mathcal{Y})\} = H(\mathcal{Y})$. Par ailleurs, comme $H(\mathcal{Y}|\mathcal{X}) = 0$ on a aussi $I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{Y})$ (formule F.7).

6. Réseaux bayésiens

Dans le réseau bayésien ci-dessous, on a nécessairement



(a) $H(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{T}) = H(\mathcal{X}) + H(\mathcal{Z}) + H(\mathcal{Y}|\mathcal{X}) + H(\mathcal{T}|\mathcal{Y}, \mathcal{Z}, \mathcal{X})$

Vrai

En toute généralité on a (règle de chaînage des entropies) $H(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{T}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) + H(\mathcal{Z}|\mathcal{X}, \mathcal{Y}) + H(\mathcal{T}|\mathcal{X}, \mathcal{Y}, \mathcal{Z})$. La structure de réseau bayésien exprime (entre autres) le fait que $P(\mathcal{Z}|\mathcal{X}, \mathcal{Y}) = P(\mathcal{Z})$ puisque \mathcal{Z} n'a pas de père et que \mathcal{X} et \mathcal{Y} sont des non-descendants de \mathcal{Z} . On a donc bien $H(\mathcal{Z}|\mathcal{X}, \mathcal{Y}) = H(\mathcal{Z})$.

(b) $I(\mathcal{V}; \mathcal{T}) \geq I(\mathcal{V}; \mathcal{X}, \mathcal{Y}, \mathcal{Z})$

Vrai

Pour le voir regroupons les variables $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ et marginalisons les variables \mathcal{U}, \mathcal{W} qui n'interviennent pas dans cette question; on obtient le réseau de la figure (b), où on voit aisément que les trois variables $\mathcal{A}, \mathcal{T}, \mathcal{V}$ forment une chaîne de Markov (dans cet ordre). La propriété est donc une conséquence de la propriété d'érosion de l'information le long d'une chaîne de Markov.

(c) $\mathcal{Y} \leftrightarrow \mathcal{U} \leftrightarrow \mathcal{W}$

Vrai

Les trois variables $\mathcal{Y}, \mathcal{U}, \mathcal{W}$ forment aussi une chaîne de Markov (voir figure (c)).

(d) $I(\mathcal{V}; \mathcal{W} | \mathcal{T}, \mathcal{U}) = 0$

Vrai

On a par définition $I(\mathcal{V}; \mathcal{W} | \mathcal{T}, \mathcal{U}) = \sum_{i,j,k,l} P(V_i, W_j, U_k, T_l) \log \frac{P(V_i, W_j | U_k, T_l)}{P(V_i | U_k, T_l) P(W_j | U_k, T_l)}$.

On a par ailleurs $(\forall i, j, k, l) P(V_i | W_j, U_k, T_l) = P(V_i | T_l) = P(V_i | U_k, T_l)$ puisque \mathcal{T} est le père de \mathcal{V} et que \mathcal{W}, \mathcal{U} en sont des non-descendants. On peut donc factoriser $P(V_i, W_j | U_k, T_l) = P(W_j | U_k, T_l) P(V_i | W_j, U_k, T_l) = P(W_j | U_k, T_l) P(V_i | U_k, T_l)$.

7. Entropies de sources

Soit $\mathcal{X}_i, \forall i > 0$ une source de Markov discrète. Lesquelles des affirmations sont toujours vraies, sous cette hypothèse ($\forall n$)

(a) $H(\mathcal{X}^n) < \sum_{i=1}^n H(\mathcal{X}_i)$

Faux

Puisqu'une source sans mémoire est un cas particulier de chaîne de Markov, on pourrait avoir l'égalité.

(b) $H(\mathcal{X}^n) = \sum_{i=1}^n H(\mathcal{X}_i | \mathcal{X}_{i-1})$

Vrai

A cause de l'hypothèse de Markov on a $H(\mathcal{X}_i | \mathcal{X}^{i-1}) = H(\mathcal{X}_i | \mathcal{X}_{i-1})$.

(c) $H(\mathcal{X}^n) = \sum_{i=1}^n H(\mathcal{X}_i | \mathcal{X}^{i-1})$

Vrai

C'est la propriété générale de chaînage des entropies.

(d) $\frac{H(\mathcal{X}^n)}{n}$ forme une suite (pas nécessairement strictement) décroissante.

Faux

Un contre-exemple peut être obtenu aisément en considérant un processus non-stationnaire, par exemple sans mémoire. On pourrait très bien avoir par exemple $H(\mathcal{X}_1) = 0$, puis $H(\mathcal{X}_i) = 1, \forall i > 1$, auquel cas on aurait $\frac{H(\mathcal{X}^n)}{n} = \frac{n-1}{n}$.

8. Ensembles typiques

On considère un processus aléatoire qui consiste à lancer successivement un pièce supposée parfaite. Soit \mathcal{X}_i le résultat du i -ème lancer, et \mathcal{X}^n le résultat des n premiers lancers.

(a) Quelle est la taille de A_n^ϵ pour ce processus ?

2^n : toutes les suites de longueur n sont équiprobables et donc forcément toutes typiques. Il y a 2^n suites de longueur n .

(b) Quelle est la probabilité de A_n^ϵ ?

1, vu ce qui précède.

(c) Quelle est le débit d'entropie du processus aléatoire \mathcal{X}_i ?

1 Shannon par symbole.

$$H(\mathcal{X}^n) = \log 2^n = n \Rightarrow \lim_{n \rightarrow \infty} \frac{H(\mathcal{X}^n)}{n} = 1.$$

(d) Comment coder de façon optimale les réalisations de ce processus ?

Un code binaire trivial ($Face \rightarrow 0, Pile \rightarrow 1$).

B. Exercices

1. Urnes et tirage sans remise

Soit une urne contenant r billes rouges et n billes noires ($n + r > 1$).

On tire deux billes successivement et sans remise, et on désigne par \mathcal{C}_1 la variable aléatoire qui représente la couleur de la première et \mathcal{C}_2 celle de la seconde.

- (a) Calculez (formellement) $H(\mathcal{C}_1)$ et $H(\mathcal{C}_2)$. Expliquez le résultat obtenu.

Solution

On obtient pour la première bille $H(\mathcal{C}_1) = -[\frac{n}{n+r} \log \frac{n}{n+r} + \frac{r}{n+r} \log \frac{r}{n+r}]$ et la même chose pour la seconde bille.

Le protocole de tirage est en fait équivalent à tirer en même temps deux billes, puis à regarder séparément la couleur des deux billes, ce qui explique la symétrie.

- (b) Calculer, en supposant que $n = 1$ et $r = 2$ les valeurs de $H(\mathcal{C}_1)$, $H(\mathcal{C}_2)$, $I(\mathcal{C}_1; \mathcal{C}_2)$.

Solution

On a $H(\mathcal{C}_1) = H(\mathcal{C}_2) = -[\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3}] = \log 3 - \frac{2}{3} = 0.9182958$.

Par ailleurs, en fonction de la couleur de la première bille $H(\mathcal{C}_2 | \mathcal{C}_1 = n) = 0$ et $H(\mathcal{C}_2 | \mathcal{C}_1 = r) = 1$. Donc $H(\mathcal{C}_2 | \mathcal{C}_1) = \frac{2}{3}$. Donc $I(\mathcal{C}_1; \mathcal{C}_2) = H(\mathcal{C}_2) - H(\mathcal{C}_2 | \mathcal{C}_1) = 0.2516291$.

- (c) On continue le tirage sans remise, jusqu'à épuisement du stock et on désigne le résultat des $n + r$ tirages successifs par $\mathcal{C}^{n+r} = \mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_{n+r}$. Calculez (formellement, et en réfléchissant) la valeur de $H(\mathcal{C}^{n+r})$.

Solution

Il suffit de voir que toutes les réalisations de \mathcal{C}^{n+r} sont équiprobables. Comme il y en a $\frac{(n+r)!}{n!r!}$ on a $H(\mathcal{C}^{n+r}) = \log \frac{(n+r)!}{n!r!}$.

Calculons la probabilité d'une suite particulière, par exemple $nn \dots nrr \dots r$. On a $P(\mathcal{C}^{n+r} = nn \dots nrr \dots r) = \frac{n}{n+r} \frac{n-1}{n+r-1} \dots \frac{1}{r+1} \frac{r}{r} \frac{r-1}{r-1} \dots \frac{1}{1} = \frac{n!r!}{(n+r)!}$. On voit que le fait de modifier l'ordre d'apparition des billes rouges et noires ne change rien à cette valeur.

2. Propriétés des codes

Pour chacun de ces codes, indiquer (en justifiant) s'il est régulier (R), déchiffrable (D), instantané (I), complet (C).

- (a) le code binaire de mots $\{0, 01, 11\}$

Régulier : puisque tous les mots de code sont différents.

Déchiffrable : si on inverse le sens de la lecture des messages encodés, on se rend compte que le code est sans préfixes et donc déchiffrable; le code de départ l'est donc aussi.

Pas Instantané : puisque le mot 0 est préfixe de 01.

Pas complet : au sens strict, car cette notion n'est définie que pour les codes instantanés. Cependant, le code "retourné" est bien un code instantané complet (il satisfait l'égalité de Kraft).

- (b) le code binaire de mots $\{00, 11, 111, 0110\}$

Régulier : oui

Déchiffrable : non, l'extension d'ordre deux du code contient deux mots identiques : 11,111 et 111,11.

Instantané : non car 11 est un préfixe de 111.

Complet : non, car pas instantané (les longueurs ne vérifient pas non plus l'égalité de Kraft).

- (c) le code ternaire de mots $\{00, 012, 011, 100, 201, 212, 22\}$

Le code est sans préfixes, donc **instantané, régulier et déchiffrable**.

Par contre, il n'est pas complet (on peut ajouter, par exemple, le mot 010 en gardant la propriété de code sans préfixes).

3. Arbres de codes, inégalité de Kraft, Huffman

Soient les longueurs de mots de codes suivantes

$$\begin{pmatrix} l_1 & l_2 & l_3 & l_4 & l_5 & l_6 & l_7 & l_8 & l_9 & l_{10} \\ 2 & 3 & 3 & 2 & 3 & 4 & 2 & 3 & 2 & 2 \end{pmatrix}.$$

- (a) Existe-t-il un code binaire déchiffrable respectant ces longueurs de mots ?

Réponse : non

En effet, calculons le membre de gauche de l'inégalité de Kraft pour $q = 2$. On a $\sum_{i=1}^{10} 2^{-l_i} = 1.8125$.

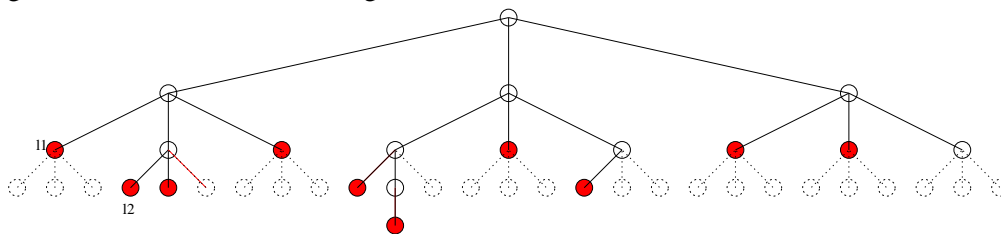
- (b) Quelle est la taille minimale de l'alphabet de code ("arité") telle que l'on puisse construire un code instantané respectant ces longueurs de mots ?

On calcule le membre de gauche de l'inégalité de Kraft pour $q = 3$. On a $\sum_{i=1}^{10} 3^{-l_i} = 0.7160493$. Il existe donc bien un code ternaire déchiffrable ayant ces longueurs de mots.

- (c) Dessiner un arbre de code (avec cet alphabet) qui réalise les longueurs données.

Réponse

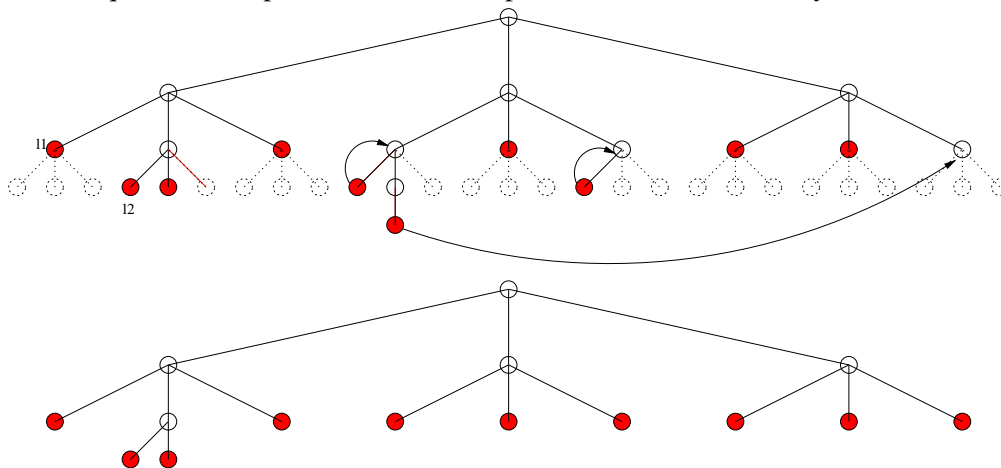
On dessine un arbre de code complet de profondeur 4 (longueur du mot le plus long) et on insère progressivement les mots de code dans cet arbre. Cela donne, si on insère les mots de gauche à droite, l'arbre de la figure ci-dessous.



- (d) Pouvez-vous modifier cet arbre (en préservant le nombre de symboles source) de façon à garantir une diminution de la longueur moyenne, quelle que soit la distribution de probabilité des symboles source ? Allez aussi loin que possible dans cette direction.

Réponse

Il faut modifier l'arbre en déplaçant les feuilles vers le haut tout en préservant le caractère sans préfixes. La figure suivante montre une telle séquence maximale de réduction de l'arbre. On voit qu'il n'existe pas d'arbre meilleur pour une distribution des symboles uniforme.



- (e) Soit, alors la distribution de probabilité suivante

$$(0.01, 0.01, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.84).$$

Associez ces probabilités aux feuilles de l'arbre obtenu à l'étape précédente de manière à minimiser la longueur moyenne de codage.

Réponse

Il suffit d'associer les probabilités les plus faibles aux feuilles les plus profondes. En l'occurrence, les probabilités 0.01 sont ici associées aux noeuds de profondeur 3.

- (f) Comparez l'entropie de la source et la longueur moyenne du code. En déduire qu'il doit exister un code pour les symboles source (utilisant le même alphabet) qui est meilleur.

Réponse

L'entropie de la source vaut ici $H = 1.1343095$ Shannon.

On en déduit qu'il existe un code ternaire de longueur moyenne inférieure à $\frac{H}{\log 3} + 1 = 1.7156696$.

La longueur moyenne du code vaut 2.02.

Il existe donc bien un code qui doit être meilleur.

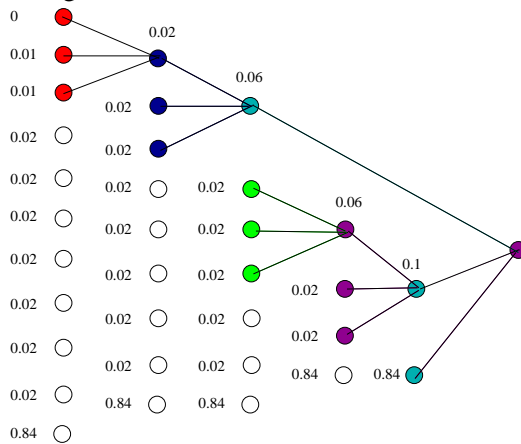
- (g) Trouvez un tel code de longueur moyenne minimale et calculez la longueur moyenne correspondante.

Réponse

Il faut appliquer l'algorithme de Huffman. Cependant, avant d'appliquer cet algorithme il faut compléter la source au moyen de symboles "bidon" de probabilité nulle, pour que le nombre de symboles sources soit compatible avec le nombre de feuilles d'un arbre ternaire complet ou incomplet ($F = 2n + 1, n \geq 0$).

Dans le cas présent, il faut ajouter un seul symbole bidon ($F=11$).

La figure ci-dessous montre la construction de l'arbre de Huffman.



La longueur moyenne du code vaut 1.24 symbole de code par symbole source. Cette valeur est bien inférieure à 1.7156696 et supérieure à $\frac{H}{\log 3} = 0.7156696$.

4. Entropies de messages codés

On s'intéresse à la relation entre l'entropie par symbole d'une source et celle de la source obtenue après codage.

Soit \mathcal{S} une variable aléatoire dont les valeurs sont $\{1, 2, 3\}$. Soit une source qui émet des séquences de symboles $\mathcal{S}^n = (\mathcal{S}_1, \dots, \mathcal{S}_n)$ indépendants et identiquement distribués comme \mathcal{S} .

D'autre part, désignons par $C(\mathcal{S})$ la séquence binaire obtenue en codant le symbole \mathcal{S} , et par $C(\mathcal{S}^n) = C(\mathcal{S}_1)C(\mathcal{S}_2) \dots C(\mathcal{S}_n)$ la séquence qui correspond au codage d'un message \mathcal{S}^n .

- (a) En supposant que C est un code régulier mais non déchiffirable, comparez
- $H(C(\mathcal{S}))$ avec $H(\mathcal{S})$,

Réponse

On a $H(C(\mathcal{S})) \leq H(\mathcal{S})$, car $C(\mathcal{S})$ est une fonction de \mathcal{S} . D'autre part, comme le code est régulier il existe aussi une fonction $g(\cdot)$, telle que $\mathcal{S} = g(C)$ et donc $H(\mathcal{S}) \leq H(C(\mathcal{S}))$. Au total $H(C(\mathcal{S})) = H(\mathcal{S})$.

ii. $H(C(\mathcal{S}^n))$ avec $H(\mathcal{S}^n)$.

Réponse

En général (quel que soit n) tout ce qu'on peut dire, c'est que $H(C(\mathcal{S}^n)) \leq H(\mathcal{S}^n)$. Mais comme le code est non déchiffrable, il existe n tel que $H(C(\mathcal{S}^n)) < H(\mathcal{S}^n)$. Par exemple, supposons que la source soit binaire et que le code associé soit

$$\begin{aligned} s_1 &\rightarrow 11 \\ s_2 &\rightarrow 111 \end{aligned}$$

Ce code est régulier. En supposant que les deux symboles source soient équiprobables, on $H(\mathcal{S}) = 1$ et $H(\mathcal{S}^2) = 2$.

D'autre part, $C(\mathcal{S}^2) = \{1111, 11111, 11111\}$ puisque les séquences s_1s_2 et s_2s_1 donnent toutes les deux le même message 11111. Les probabilités associées à la variable $C(\mathcal{S}^2)$ sont $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\}$ et donc l'entropie de cette variable vaut $H(C(\mathcal{S}^2)) = \frac{3}{2}$.

(b) En supposant que la loi de probabilité associée à \mathcal{S} est

$$P(1) = 1/2; P(2) = P(3) = 1/4,$$

i. trouvez l'entropie par symbole de source (en bits par symbole de source).

Réponse : $\lim_{n \rightarrow \infty} \frac{H(\mathcal{S}^n)}{n} = H(\mathcal{S}_1) = -[\frac{1}{2} \log \frac{1}{2} + \frac{1}{4} \log \frac{1}{4} + \frac{1}{4} \log \frac{1}{4}] = \frac{3}{2}$ Shannon par symbole.

ii. Soit le code

$$C(1) = 0, C(2) = 10, C(3) = 11.$$

Trouvez l'entropie de la source codée (en bits par symbole binaire). Remarquez que la source codée n'est plus compressible.

Réponse

Intuitivement : le code est déchiffrable et d'autre part la longueur moyenne du code est égale à l'entropie de la source. Il s'agit donc d'un code optimal, et les messages une fois encodés doivent donc former une suite de symboles successifs indépendants et équiprobables (car sinon on pourrait les compresser encore, et donc le code pour \mathcal{S} ne serait pas optimal. Le débit d'entropie est donc de 1 Shannon par bit.

iii. Soit maintenant le code suivant (de longueurs uniformes)

$$C(1) = 00, C(2) = 10, C(3) = 01.$$

Que devient l'entropie par symbole (binaire) de la source codée ?

Réponse

Intuitivement : le code est encore déchiffrable et produit exactement 2 bits par symbole source. Comme la source de départ a un débit d'entropie de 1.5 Shannon par symbole, la source binaire résultant de l'encodage produit 0.75 Shannon par bit.

Plus rigoureusement : il s'agit de déterminer si la suite $\frac{H(C^n)}{n}$ converge et vers quelle limite. Or cette suite est coincée entre les deux suites suivantes

$$\frac{H(\mathcal{S}^{\lfloor n/2 \rfloor})}{n} \leq \frac{H(C^n)}{n} \leq \frac{H(\mathcal{S}^{\lceil n/2 \rceil})}{n}.$$

Comme ces deux suites convergent vers une même limite à savoir $\frac{H(\mathcal{S})}{2} = \frac{3}{4}$ il en est de même pour $\frac{H(C^n)}{n}$.

Formules utiles

Entropie, information, divergence

$$H(\mathcal{X}) \triangleq - \sum_{i=1}^n P(X_i) \log P(X_i) \quad (\text{F.1})$$

$$H(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m) \triangleq - \sum_{i_1=1}^{n_1} \dots \sum_{i_m=1}^{n_m} P(X_{1,i_1}, \dots, X_{m,i_m}) \log P(X_{1,i_1}, \dots, X_{m,i_m}) \quad (\text{F.2})$$

$$H(\mathcal{X}|Y_j) = - \sum_{i=1}^n P(X_i|Y_j) \log P(X_i|Y_j) \quad (\text{F.3})$$

$$H(\mathcal{X}|\mathcal{Y}) \triangleq - \sum_{i=1}^n \sum_{j=1}^m P(X_i, Y_j) \log P(X_i|Y_j) \quad (\text{F.4})$$

$$H(\mathcal{X}|\mathcal{Y}) = \sum_{j=1}^m P(Y_j) H(\mathcal{X}|Y_j) \quad (\text{F.5})$$

$$I(\mathcal{X}; \mathcal{Y}) \triangleq \sum_{i=1}^n \sum_{j=1}^m P(X_i, Y_j) \log \frac{P(X_i, Y_j)}{P(X_i)P(Y_j)} \quad (\text{F.6})$$

$$I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) \quad (\text{F.7})$$

$$= H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X}, \mathcal{Y}) \quad (\text{F.8})$$

$$I(\mathcal{X}; \mathcal{Y}|\mathcal{Z}) \triangleq H(\mathcal{X}|\mathcal{Z}) - H(\mathcal{X}|\mathcal{Y}, \mathcal{Z}) \quad (\text{F.9})$$

$$I(\mathcal{X}; \mathcal{Y}|\mathcal{Z}) = \sum_{i,j,k} P(X_i, Y_j, Z_k) \log \frac{P(X_i, Y_j|Z_k)}{P(X_i|Z_k)P(Y_j|Z_k)} \quad (\text{F.10})$$

$$D(P||Q) \triangleq \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)} \quad (\text{F.11})$$

Additivité et chaînage

$$H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{Y}) + H(\mathcal{X}|\mathcal{Y}) \quad (\text{F.12})$$

$$H(\mathcal{X}, \mathcal{Y}|\mathcal{Z}) = H(\mathcal{X}|\mathcal{Z}) + H(\mathcal{Y}|\mathcal{X}, \mathcal{Z}) \quad (\text{F.13})$$

$$\begin{aligned} H(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n) &= H(\mathcal{X}_1) + H(\mathcal{X}_2|\mathcal{X}_1) + H(\mathcal{X}_3|\mathcal{X}_2, \mathcal{X}_1) + \dots + H(\mathcal{X}_n|\mathcal{X}_{n-1}, \dots, \mathcal{X}_1) \\ &\triangleq \sum_{i=1}^n H(\mathcal{X}_i|\mathcal{X}_{i-1}, \dots, \mathcal{X}_1) \end{aligned} \quad (\text{F.14})$$

$$\begin{aligned} H(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n|\mathcal{Y}) &= H(\mathcal{X}_1|\mathcal{Y}) + H(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}) + \dots + H(\mathcal{X}_n|\mathcal{X}_{n-1}, \dots, \mathcal{X}_1, \mathcal{Y}) \\ &\triangleq \sum_{i=1}^n H(\mathcal{X}_i|\mathcal{X}_{i-1}, \dots, \mathcal{X}_1, \mathcal{Y}) \end{aligned} \quad (\text{F.15})$$

$$I(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n; \mathcal{Y}) = \sum_{i=1}^n I(\mathcal{X}_i; \mathcal{Y}|\mathcal{X}_{i-1}, \dots, \mathcal{X}_1) \quad (\text{F.16})$$

Positivité, convexité et monotonie

$$H(\mathcal{X}) \geq 0 \quad \text{et} \quad H(\mathcal{X}, \mathcal{Y}) \geq 0 \quad \text{et} \quad H(\mathcal{X}|\mathcal{Y}) \geq 0 \quad (\text{F.17})$$

$$H(\mathcal{X}, \mathcal{Y}) \leq H(\mathcal{X}) + H(\mathcal{Y}) \leq 2H(\mathcal{X}, \mathcal{Y}) \quad (\text{F.18})$$

$$0 \leq H(\mathcal{X}) \leq H(\mathcal{X}, \mathcal{Y}) \leq H(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) \leq \dots \quad (\text{F.19})$$

$$H(\mathcal{X}) \geq H(\mathcal{X}|\mathcal{Y}) \geq H(\mathcal{X}|\mathcal{Y}, \mathcal{Z}) \geq \dots \geq 0 \quad (\text{F.20})$$

$$I(\mathcal{X}; \mathcal{Y}) \geq 0 \quad \text{et} \quad I(\mathcal{X}; \mathcal{Y}|\mathcal{Z}) \geq 0 \quad (\text{F.21})$$

$$0 \leq I(\mathcal{X}; \mathcal{Y}) \leq I(\mathcal{X}; \mathcal{Y}, \mathcal{Z}) \leq I(\mathcal{X}, \mathcal{T}; \mathcal{Y}, \mathcal{Z}) \quad (\text{F.22})$$

$$D(P||Q) \geq 0 \quad (\text{F.23})$$

Non-cr ation d'information

Si $\mathcal{X} \leftrightarrow \mathcal{Y} \leftrightarrow \mathcal{Z}$ forment une cha ne de Markov alors

$$I(\mathcal{X}; \mathcal{Y}) \geq I(\mathcal{X}; \mathcal{Z}) \quad \text{et} \quad I(\mathcal{Y}; \mathcal{Z}) \geq I(\mathcal{X}; \mathcal{Z}) \quad (\text{F.24})$$

En particulier $\mathcal{Z} = g(\mathcal{Y})$

$$I(\mathcal{X}; \mathcal{Y}) \geq I(\mathcal{X}; g(\mathcal{Y})) \quad (\text{F.25})$$

In galit  de Kraft

Soient n_1, \dots, n_Q des longueurs de mots candidates pour coder une source Q -aire dans un alphabet q -aire. Alors l'in galit  de Kraft

$$\sum_{i=1}^Q q^{-n_i} \leq 1 \quad (\text{F.26})$$

est une condition n cessaire et suffisante d'existence d'un code d chiffable respectant ces longueurs de mots.