Théorie de l'information et du codage - Examen écrit
Partie Théorie - 19 janvier 2005

La majorité des questions ci-dessous sont du type "vrai/faux" : il s'agit d'indiquer si les affirmations énoncées sont vraies ou fausses. Pour les questions marquées d'une astérisque (*), on demande de justifier la réponse : si la réponse est "vrai", justifier par une propriété ou un développement; si la réponse est "faux", donner un contre-exemple.

D'autres questions demandent une réponse ouverte et peuvent nécessiter quelques développements élémentaires.

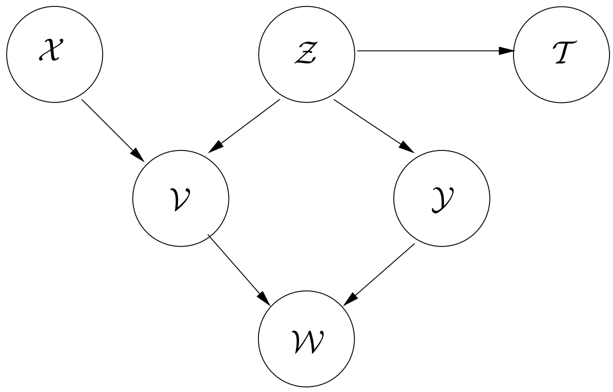
1. Indépendances d'événements
   Soient A, B et C trois événements, sous-ensembles d'un univers Ω.
   (a) $A \perp B \Leftrightarrow A \perp \neg B$
   (b) $\forall A\,,\ A \perp \Omega$
   (c) $A \perp B \Rightarrow P(A \cup B) = P(A) + P(B)$
   (d) (*) $\forall B$ tel que $P(B) \neq 0,\ P(B) \neq 1\,,\ A \perp B \Leftrightarrow P(A|B) = P(A|\neg B)$
2. Entropies
   (a) $H(\mathcal{X}|\mathcal{Y}) = \log_2 |\mathcal{X}| \Leftrightarrow \mathcal{X} \perp \mathcal{Y}$
   (b) $H(\mathcal{X},\mathcal{Y}) - H(\mathcal{X}|\mathcal{Z}) \geq H(\mathcal{X},\mathcal{Y}) - H(\mathcal{X}|\mathcal{Y},\mathcal{Z})$
   (c) $H(\mathcal{X},\mathcal{Y},\mathcal{Z},\mathcal{T}) = H(\mathcal{T}) + H(\mathcal{Z}|\mathcal{T}) + H(\mathcal{X},\mathcal{Y}|\mathcal{Z},\mathcal{T})$
   (d) (*) $H(\mathcal{X}|\mathcal{Y},\mathcal{Z}) = 0 \Rightarrow$ soit $H(\mathcal{X}|\mathcal{Y})$, soit $H(\mathcal{X}|\mathcal{Z}) = 0$
   (e) $\mathcal{X} \perp \mathcal{Y} \Rightarrow H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{Y}|\mathcal{X})$
3. Informations mutuelles
   (a) $P(\mathcal{U},\mathcal{X},\mathcal{Y}) = P(\mathcal{U},\mathcal{X})P(\mathcal{Y}|\mathcal{X}) \Rightarrow I(\mathcal{U};\mathcal{Y}) \leq I(\mathcal{X};\mathcal{Y})$
   (b) (*) $I(\mathcal{X};\mathcal{Y}) = 0$ et $I(\mathcal{X};\mathcal{Z}) = 0 \Rightarrow I(\mathcal{X};\mathcal{Y},\mathcal{Z}) = 0$
   (c) $I(\mathcal{X};\mathcal{Z}|\mathcal{Y}) = I(\mathcal{Z};\mathcal{Y}|\mathcal{X}) - I(\mathcal{Z};\mathcal{Y}) + I(\mathcal{X};\mathcal{Z})$
   (d) $\mathcal{X} \perp \mathcal{Z}|\mathcal{Y} \Rightarrow I(\mathcal{X};\mathcal{Y},\mathcal{Z}) = I(\mathcal{X};\mathcal{Y})$
4. Fonctions de variables aléatoires
   (a) $\mathcal{Y} = f(\mathcal{X}) \Rightarrow I(\mathcal{X};\mathcal{Y}) = H(\mathcal{Y})$
   (b) $H(\mathcal{X}, g(\mathcal{X})) = H(\mathcal{X})$
   (c) $H(\mathcal{X}, g(\mathcal{X})) > H(g(\mathcal{X}))$
   (d) (*) $H(\mathcal{X}|g(\mathcal{Y})) = H(\mathcal{X}|\mathcal{Y}) \Rightarrow \mathcal{X} \perp \mathcal{Y}|g(\mathcal{Y})$

5. Factorisations de distributions conjointes
   Soit $P(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}) = P(\mathcal{A})P(\mathcal{B})P(\mathcal{C}|\mathcal{B})P(\mathcal{D}|\mathcal{B})$.

   On a nécessairement

   (a) $P(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}) = P(\mathcal{C})P(\mathcal{A})P(\mathcal{B}|\mathcal{C})P(\mathcal{D}|\mathcal{B})$
   (b) $P(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}) = P(\mathcal{C})P(\mathcal{A}|\mathcal{C})P(\mathcal{B}|\mathcal{C})P(\mathcal{D}|\mathcal{C})$
   (c) (*) $P(\mathcal{A},\mathcal{B},\mathcal{C}) = P(\mathcal{A})P(\mathcal{B})P(\mathcal{C}|\mathcal{B})$
   (d) $P(\mathcal{A},\mathcal{C},\mathcal{D}) = P(\mathcal{A})P(\mathcal{C})P(\mathcal{D}|\mathcal{C})$
   (e) $P(\mathcal{B},\mathcal{C},\mathcal{D}) = P(\mathcal{B})P(\mathcal{C})P(\mathcal{D}|\mathcal{C})$

6. Réseaux bayésiens et indépendances conditionnelles
   Dans le réseau bayésien ci-dessous, on a nécessairement



   (a) (*) $\mathcal{Z} \leftrightarrow \mathcal{Y} \leftrightarrow \mathcal{W}$
   (b) $I(\mathcal{X};\mathcal{W}) \geq I(\mathcal{X};\mathcal{Z},\mathcal{Y},\mathcal{T})$
   (c) $H(\mathcal{Z},\mathcal{Y},\mathcal{V},\mathcal{W}) = H(\mathcal{W},\mathcal{V}) + H(\mathcal{Y}|\mathcal{V},\mathcal{W}) + H(\mathcal{Z}|\mathcal{V},\mathcal{W})$
   (d) $\mathcal{Z} \perp \mathcal{T}|\mathcal{T}$

7. Sources d'information
   On considère les suites suivantes de variables aléatoires binaires :

$$\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_n$$
$$\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_n$$
$$\mathcal{W}_0, \mathcal{W}_1, \dots, \mathcal{W}_n$$

   dont les distributions conjointes se factorisent comme suit :

$$P(\mathcal{X}_0,\dots,\mathcal{X}_n,\mathcal{Y}_0,\dots,\mathcal{Y}_n) = P(\mathcal{Y}_0|\mathcal{X}_0)\prod_{i=0}^{n} P(\mathcal{X}_i)\prod_{k=1}^{n} P(\mathcal{Y}_k|\mathcal{X}_k,\mathcal{Y}_{k-1})$$
$$P(\mathcal{W}_0,\dots,\mathcal{W}_n) = \prod_{i=0}^{n} P(\mathcal{W}_i)$$

   Trois sources S_1, S_2 et S_3 sont modélisées par, respectivement, les suites de variables aléatoires

$$\mathcal{X}_0, \mathcal{X}_1, \dots$$
$$\mathcal{Y}_0, \mathcal{Y}_1, \dots$$
$$\mathcal{Z}_0, \mathcal{Z}_1, \dots$$

   avec

$$\mathcal{Z}_i = \mathcal{Y}_i + \mathcal{W}_i$$

(a) S_1 est une source sans mémoire

(b) S_3 est une source sans mémoire

(c) S_3 est une source de Markov d'ordre 1

(d) S_2 est une source de Markov d'ordre 1

(e) $\mathcal{Z}_i \perp \mathcal{Z}_k | \mathcal{Y}_j$, $1 \leq i < n$, $1 < k \leq n$, $i \leq j \leq k$

(f) $\mathcal{Z}_i \not\perp \mathcal{Z}_k | \mathcal{Y}_j$, $1 \leq i < n$, $1 < k \leq n$, $j < i$, $j > k$

8. Propriétés de sources
Soit $\mathcal{X}_i$ une suite de variables indépendantes et identiquement distribuées selon :

$$P(\mathcal{X}_i = 1) = p \,, \quad P(\mathcal{X}_i = -1) = 1 - p \quad i = 1, 2, \ldots$$

On déduit de cette suite la suite de variables $\mathcal{Z}_k$ selon :

$$\mathcal{Z}_k = \sum_{i=1}^{k} \mathcal{X}_i$$

(a) Les valeurs possibles de $\mathcal{Z}$ sont $\{-1, 0, 1\}$

(b) $\mathcal{Z}$ est un processus stationnaire

(c) $\mathcal{Z}$ est un processus ergodique

(d) $\mathcal{Z}$ est un processus de Markov (d'ordre 1)

(e) $H(\mathcal{Z}_1, \ldots, \mathcal{Z}_n) = H(\mathcal{X}_1, \ldots, \mathcal{X}_n)$

(f) (*) Le processus $\mathcal{Z}$ possède un débit d'entropie bien défini.

9. Entropies de sources
Soit $\mathcal{X}_i, \ \forall i > 0$, une source de Markov discrète (d'ordre 1). On note $\mathcal{X}^n$ la suite de variables $\mathcal{X}_1, \ldots, \mathcal{X}_n$.

(a) $H(\mathcal{X}^n) = H(\mathcal{X}_1) + \sum_{i=2}^{n} H(\mathcal{X}_i | \mathcal{X}_{i-1})$

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

(c) $H(\mathcal{X}^n) = H(\mathcal{X}_1) + (n-1) H(\mathcal{X}_n | \mathcal{X}_{n-1})$

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

10. Codage de sources
Le code ternaire $\{0, 20, 021\}$ est :

(a) régulier

(b) (*) déchiffrable

(c) un code de Huffman

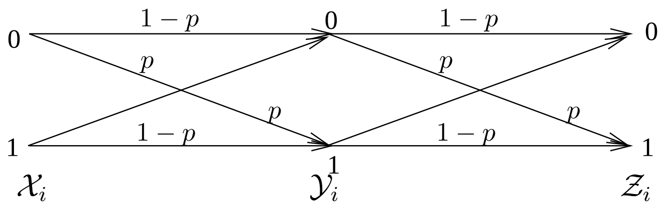
(d) instantané

(e) complet

11. Codes de Huffman
Lesquels, parmi les ensembles suivants de longueurs de mots de codes, peuvent correspondre à un code de Huffman ternaire ? Justifiez chaque réponse.

(a) (1,1,2,2,3,3,3)
(b) (1,1,2,2,3,3)
(c) (1,1,2,2,3)
(d) (1,2,2,2,2,2,2)
(e) (1,2,2,2,2)

12. Propriétés des canaux discrets
Soit un canal sans mémoire binaire constitué par la mise en cascade de deux canaux symétriques binaires de même probabilité d'erreur p.



Soient :

- $\mathcal{X}_i$ le processus aléatoire d'entrée
- $\mathcal{Y}_i$ le processus de sortie du premier canal
- $\mathcal{Z}_i$ la processus de sortie du second canal

(a) Que vaut la capacité en information C_1 de chacun des deux canaux constitutifs ?
(b) Le canal complet résultant de la mise en cascade est un canal symétrique binaire
(c) Une suite de symboles indépendants $\mathcal{X}_i$ en entrée conduit à une suite de symboles indépendants $\mathcal{Z}_i$ en sortie
(d) Si les entrées $\mathcal{X}_i$ sont équiprobables, les sorties $\mathcal{Z}_i$ le sont aussi
(e) La capacité en information du canal complet C_2 est maximale lorsque $p = \frac{1}{4}$
(f) (*) On a toujours $C_2 < C_1$

13. Propriétés des canaux discrets

    Soit un canal à bruit additif avec 2 niveaux d'entrée et de sortie ($\mathcal{X} = \{0, 1\} = \mathcal{Y}$). Désignons par $\mathcal{X}_i, \mathcal{Y}_i, \mathcal{Z}_i$ les processus aléatoires d'entrée, de sortie et de bruit. Le processus de bruit est de Markov d'ordre 1 caractérisé par la loi de probabilité $P(Z_0 = 0) = P(Z_0 = 1) = 0.5$ et la condition $P(Z_{i+1} = Z_i) = (1 - p)$. La relation entrée-sortie est $Y_i = (X_i + Z_i)$mod2. Soit $C(p)$ la capacité en information du canal.

    (a) Le processus de bruit est stationnaire et ergodique
    (b) (*) Le processus de sortie est un processus de Markov
    (c) Si le processus d'entrée est sans mémoire, le processus de sortie est sans mémoire
    (d) Réciproquement, si le processus de sortie est sans mémoire, le processus d'entrée est sans mémoire
    (e) Si les entrées sont équiprobables, les sorties sont équiprobables
    (f) (*) Si les entrées sont indépendantes et équiprobables, les sorties le sont aussi
    (g) La capacité du canal est atteinte lorsque les entrées sont équiprobables et indépendantes
    (h) $H(X^n|Y^n) = H(Z^n)$
    (i) Que vaut la capacité en information de ce canal ?

14. Codage de canal

    Soit le code de canal $(n = 5, M = 8)$ suivant

$$C = \begin{array}{c} 00000 \\ 00101 \\ 01011 \\ 01110 \\ 10001 \\ 10100 \\ 11010 \\ 11111 \end{array}$$

    (a) (*) Le code est systématique
    (b) Le code est déchiffrable
    (c) Le code est linéaire
    (d) (*) Le code permet la détection d'erreurs simples
    (e) Le débit du code vaut ?
    (f) Le message 010 est envoyé sur le canal, et en sortie on lit la suite 00001. Le décodeur se trompera-t-il ?