Examen de théorie de l'information - Ensimag - 1A

Jeudi 28 janvier 2016, 14h-16h.

POLYCOPIE, NOTES DE COURS ET CALCULATRICE AUTORISES

Le sujet est composé de 4 problèmes indépendants. Le barème est donné à titre indicatif.

1 CODAGE SOURCE (5 points)

1. On considère une suite de variables aléatoires binaires dont chaque variable est d'entropie $H = 1$. Dans quel cas l'utilisation d'un codeur de source pour compresser cette suite
   (a) présente un intérêt ?
   (b) ne présente aucun intérêt ?
2. Soit X une source simple à cinq états de probabilités $1/4$, $1/4$, $1/4$, $1/8$, $1/8$
   (a) Quelle est l'entropie de X ?
   (b) On code cette source avec un code C dont les mots sont 00, 10, 11, 101, 111, quel est la longueur moyenne des mots ?
   (c) Calculer l'efficacité du code C.
   (d) C est-il un code opérationnel ?
   (e) Construire un code de Huffman pour X.
   (f) Coder les extensions d'ordre deux de X avec un code de Huffman peut-il présenter un intérêt. Justifier.

2 CODAGE CANAL (6 points)

On considère le code bloc linéaire binaire de matrice génératrice :

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

1. Donner la taille n des mots du code, le nombre k de bits d'information.
2. Dire de combien de mots le code est composé et écrire l'ensemble des mots.
3. Calculer la distance minimale de ce code et en déduire sa capacité de correction d'erreur.

4. Ecrire la matrice de contrôle de parité $\mathbf{H}$.
5. On note $\mathbf{c}$ le mot de code en entrée du canal. Le canal est supposé binaire symétrique de probabilité de transition p. On note $\mathbf{y}$ la séquence associée en sortie du canal.
   (a) En supposant qu'il s'est produit une seule erreur, dire quelles sont les valeurs possibles du syndrome et donner un algorithme de correction d'erreur (autre que la recherche exhaustive).
   (b) Le syndrome vaut 110, que peut-on dire ?
   (c) Calculer la probabilité pour que le nombre d'erreurs dans la séquence $\mathbf{y}$ soit strictement supérieur à 1 et préciser, en le justifiant, si ce résultat coïncide avec la probabilité d'erreur par mot.

3 CASCADE DE CANAUX BINAIRES (3 points)

On met deux canaux binaires à la queue leu leu (la sortie du premier sert d'entrée au second). Leurs probabilités de transition sont respectivement p et q.

1. Montrer que la matrice de transition de cette cascade de deux canaux est le produit des deux matrices de transition.
2. Donner l'expression de la capacité du canal global.

4 OUTILS GÉNÉRAUX (6 points)

Soit un jeu de 4 cartes portant les valeurs 1 ; 2 ; 3 ; 4. On tire successivement et sans remise 2 cartes de ce jeu. On considère les trois variables aléatoires suivantes :

— X, la valeur de la première carte ;
— Y, la valeur de la deuxième carte ;
— $S = X + Y$, la somme des valeurs associées aux deux cartes.

Dans la suite, vous préciserez les distributions de probabilités ou tableaux de probabilités conditionnelles et conjointes nécessaires.

1. Montrer que l'entropie de X vaut $H(X) = 2$ bits,
2. Calculer $H(Y|X)$ et $H(X;Y)$,
3. En déduire que $H(Y) = H(X)$,
4. Calculer l'information mutuelle $I(X;Y)$,
5. Calculer l'entropie $H(S)$,
6. Calculer les entropies $H(S|X)$ et $H(X;S)$,
7. Calculer l'information mutuelle $I(X;S)$ et $H(X|S)$.

Examen de théorie de l'information - Ensimag - 1A

Novembre 2018

POLYCOPIE, NOTES DE COURS ET CALCULATRICES AUTORISES

Le sujet est composé de 5 problèmes indépendants. Le barème est donné à titre indicatif.

1 Questions générales (3 points)

Soient

X une variable aléatoire à $N = 2$ états (x_1 et x_2),

Y une variable aléatoire à $M = 3$ états (y_1, y_2 et y_3)

avec les probabilités conjointes $P(X = x_i\ ; Y = y_j)$, $i = 1, 2$; $j = 1, 2, 3$ données ci-dessous :

$P(x_i\ ; y_j)$	y_1	y_2	y_3
x_1	$1/8$	$1/8$	$1/4$
x_2	$3/8$	$1/8$	0

1. Déterminer les entropies $H(X)$ (0.5 point), $H(Y)$ (0.5 point) et $H(Y|X)$ (0.5 point) ainsi que l'information mutuelle $I(X\ ; Y)$ (0.5 point) (on donnera les valeurs exactes — avec $\log$ (0.5 point) — et approchées).
2. En déduire $H(X|Y)$. (0.5 point)

2 Capacité de canal (3 points)

On considère un canal symétrique, tel que

+ l'entrée X peut prendre $N = 3$ états notés $\{0; 1; 2\}$,

+ la sortie Y peut prendre $M = 3$ états notés $\{0; 1; 2\}$,

+ $p(y = 1|x = 0) = p$, $p(y = 2|x = 0) = q$ et $p(y = 0|x = 1) = q$.

1. Ecrire la matrice de transition de ce canal. (1 point)
2. Quelle loi d'entrée maximise l'information mutuelle entre l'entrée et la sortie ? (1 point)
3. Calculer la capacité de ce canal (1 point)

3 Déchiffrabilité (6 points)

On code une source simple à 4 états par les mots binaires

$M_1 = 1, M_2 = 101, M_3 = 10, M_4 = 01$ de longueurs $l_1 = 1$, $l_2 = 3$, $l_3 = 2$, $l_4 = 2$.

1. Montrer que ce code n'est pas instantané. (0,5 point)
2. Ce code est-il déchiffrable ? Justifier votre réponse. (0,5 point)
3. Montrer qu'il n'existe pas de code instantané binaire avec le même jeu de longueurs. (1 point)
4. Existe-t-il un code instantané ternaire ? Justifier dans le cas négatif ou donner un tel code dans le cas positif en précisant si le code donné peut-être un code de Huffman. (1,5 point)
5. A partir de maintenant, les probabilités des 4 états sont $1/2$, $1/4$, $1/8$, $1/8$. (1 point)
   Construire un code binaire instantané de longueur moyenne aussi faible que possible pour le codage de cette source.
6. Calculer l'entropie de la source. (0,5 point)
7. Montrer qu'il n'est pas nécessaire de coder des extensions de cette source pour atteindre l'efficacité maximale. (1 point)

4 Codage canal (5 points)

On considère le code bloc linéaire binaire de matrice de contrôle de parité :

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

1. Quel est le rendement de ce code ? (0,5 point)
2. Donner sa matrice génératrice. (0,5 point)
3. Dire de combien de mots le code est composé (0,5 point) et écrire l'ensemble des mots (0,5 point).
4. Calculer sa capacité de correction d'erreur. (1 point)
5. On note $\mathbf{c}$ le mot de code en entrée du canal. Le canal est supposé binaire symétrique de probabilité de transition $p < 1/2$. On note $\mathbf{y}$ la séquence observée en sortie du canal.
   (a) En supposant qu'il s'est produit une seule erreur, dire quelles sont les valeurs possibles du syndrome et donner un algorithme de correction d'erreur (autre que la recherche exhaustive). (1 point)
   (b) Le syndrome vaut 110, que peut-on dire ? (1 point)

5 Réduction de l'entropie par fusion d'états (3 points)

Soit X une v.a. discrète à valeurs dans $\{x_1, x_2, \cdots, x_{k+1}\}$ et Y la v.a. à valeurs dans

$\{y_1 = x_1, \cdots, y_{k-1} = x_{k-1}, y_k = \{x_k \cup x_{k+1}\}\}$

(C'est-à-dire que l'état y_k est obtenu en groupant les états x_k et x_{k+1} de la v.a. X.)

1. Montrer que l'entropie de Y est inférieure ou égale à celle de X. (2 point)
2. Dans quel cas a-t-on égalité ? (1 point)

Examen de THÉORIE DE L'INFORMATION - Ensimag - 1A
Polycopie, notes de cours et calculatrice autorisés

Le sujet est composé de 4 problèmes indépendants. Le barème est donné à titre indicatif.

1 CODAGE SOURCE (4 points)

On considère une source simple S à 5 états notés 'a', 'b', 'c', 'd', 'e' de probabilités $p(a) = 0,04$; $p(b) = 0,06$; $p(c) = 0,4$; $p(d) = 0,2$; $p(e) = 0,3$

1. Quelle est la redondance de cette source S ?
2. Construire C_2 un code de Fano-Shannon pour cette source.
3. Quelle est l'efficacité de C_2 ?
4. Construire C_3 un code de Huffman pour cette source.
5. Quelle est l'efficacité de C_3 ?

2 CODAGE CANAL (6 points)

On considère le code bloc linéaire binaire de matrice génératrice :

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

1. Donner la taille n des mots du code, le nombre k de bits d'information.
2. Dire de combien de mots le code est composé et écrire l'ensemble des mots.
3. Calculer la distance minimale de ce code et en déduire sa capacité de correction d'erreur.
4. Ecrire la matrice de contrôle de parité $\mathbf{H}$.
5. On note $\mathbf{c}$ le mot de code en entrée du canal. Le canal est supposé binaire symétrique de probabilité de transition p. On note $\mathbf{y}$ la séquence associée en sortie du canal.
   (a) En supposant qu'il s'est produit une seule erreur, dire quelles sont les valeurs possibles du syndrome et donner un algorithme de correction d'erreur (autre que la recherche exhaustive).
   (b) Le syndrome vaut 0101, que peut-on dire ?
   (c) Peut-on construire un code à répétition de même rendement et de même distance minimale ? Justifier en cas de réponse négative ou donner sa matrice génératrice dans le cas positif.

3 CANAL À ENTRÉE BINAIRE ET BRUIT TERNAIRE (6 points)

On considère un canal sans mémoire d'entrée $X \in \{-1, +1\}$ binaire et à bruit B additif ternaire à valeurs dans $\{-1, 0; +1\}$ avec les probabilités respectives $\{1/4; 1/2; 1/4\}$. On note $Y = X + B$ la sortie du canal.

1. Donner la matrice de transition de ce canal.
2. Calculer l'entropie conditionnelle $H(Y|X)$.
3. Est-il possible d'atteindre une loi uniforme en sortie de canal ? Dire pourquoi en quelques mots.
4. Quelle est l'entropie maximale de la sortie ? Pour quelle loi d'entrée est-elle atteinte ?
5. En déduire la capacité du canal.

4 OUTILS GÉNÉRAUX - DIVERGENCE DE KULLBACK-LEIBLER [4 points]

Soient 2 sources simples S_A et S_B de même alphabet binaire $\mathcal{A} = \{0; 1\}$ mais de lois de probabilités respectives différentes $P_{S_A} = \{1/2; 1/2\}$ et $P_{S_B} = \{3/4; 1/4\}$.

Un processus aléatoire tire au hasard la source S parmi les 2 possibles de manière équiprobable ($Pr(S = S_A) = Pr(S = S_B) = 1/2$), puis émet une succession indépendante de K variables binaires de la source retenue.

A partir de l'observation (ou réalisation) de la séquence émise $S_{\textsf{obs}} = [s_1, s_2, \cdots, s_K]$, on doit décider la source émettrice (décision notée $\hat{S}$).

1. On suppose l'observation $S_{\textsf{obs}} = [0, 1, 0, 0, 1, 0, 1, 0]$ avec $K = 8$.
   La distribution empirique (fréquences) des 0 et 1 est donc $P_{\textsf{obs}} = \{5/8; 3/8\}$.
   (a) Quelle serait la décision obtenue en utilisant comme critère la distance euclidienne entre les histogrammes empiriques ($P_{\textsf{obs}}$) et théoriques (P_{S_A} et P_{S_B}) ?
   (b) On rappelle que la règle de décision optimale au sens de la probabilité d'erreur minimale est : on décide $\hat{S} = S_A$ si $Pr(S_{\textsf{obs}}|S = S_A) > Pr(S_{\textsf{obs}}|S = S_B)$, et on décide $\hat{S} = S_B$ sinon.
   Pour $S_{\textsf{obs}} = [0, 1, 0, 0, 1, 0, 1, 0]$, calculer les 2 probabilités $Pr(S_{\textsf{obs}}|S = S_A)$ et $Pr(S_{\textsf{obs}}|S = S_B)$ et prendre la décision.
   (c) Démontrer formellement que la règle de décision optimale est équivalente à une règle de comparaison de divergences de Kullback-Leibler à préciser.
   (d) Calculer $D_{KL}(P_{\textsf{obs}}; P_{S_A})$ et $D_{KL}(P_{\textsf{obs}}; P_{S_B})$ où $D_{\textsf{KL}}$ est la divergence de Kullback-Leibler. En déduire la décision à prendre.