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

Vendredi 27 janvier 2017, 09h-11h.

POLYCOPIE, NOTES DE COURS ET CALCULATRICE AUTORISES
Le sujet est composé de 3 problèmes indépendants. Le barème est donné à titre indicatif.

Exercice I : Codage de source instantané (questions indépendantes) (7 pts)

*Soit une source simple S d'alphabet de taille **N=6 lettres**.*
*On veut appliquer à cette source un codage de source **instantané d'alphabet de taille D** (entier supérieur à 1, par exemple code binaire si D=2, code ternaire si D=3, ...).*

1) **Distribution fixée** (4 points) : on suppose ici que la distribution de probabilités des lettres de la source S est **Ps = { 2/24 ; 3/24 ; 3/24 ; 3/24 ; 4/24 ; 9/24 } :**
- **a)** Calculer l'entropie et la redondance de la source,
- **b)** Calculer la longueur moyenne minimale des mot-codes selon le 1° théorème de Shannon si D = 2, puis si D = 3,
- **c)** Ici D=2 : donner le codage binaire de Huffman de S, la longueur moyenne des mot-codes, et l'efficacité du code.

2) **Code fixé** (2 points) :
- **a)** On suppose le code binaire **C1**= {000, 001, 010, 011, 100, 101}.
  Montrer que ce code est déchiffrable.
  Existe-t-il une distribution Ps pour laquelle ce code est un code de Huffman (justifier) ?
- **b)** On suppose le code binaire **C2**= {00, 01, 100, 101, 110, 111}.
  Ce code est-il instantané ?
  Existe-t-il une distribution Ps pour laquelle ce code est 100% efficace (la donner alors) ?

3) **Longueurs des mot-codes fixées** (1 point) : indiquer (justifier) la taille minimale de l'alphabet de codage *D* permettant d'avoir : **1 mot-code de longueur 1 et 5 mot-codes de longueur 2**.

Exercice II : Information Mutuelle et Capacité (8 points)

Un vote a lieu au sein d'un club composée de 2 catégories de **sportifs** ***(S)*** ***en égale proportion*** ***(p=1/2)*** *: ceux inscrits en* **compétition** ***(c)*** *et ceux évoluant en* **loisir** ***(l)***.
*On s'intéresse à l'information mutuelle entre la catégorie de sportif **(S)** et les résultats des votes du premier tour (V_1) puis du deuxième tour (V_2).*
*La situation peut s'interpréter en terme de propagation de l'information **S** au travers de la cascade de deux canaux de transmission :* $S \rightarrow V_1 \rightarrow V_2$

1) **Premier tour de vote** (2,5 points) : *chez les sportifs « compétiteurs » **(c)**, **6/8**ème votent* **oui** ***(o)***, ***1/8**ème votent* **non** ***(n)*** *et les **1/8**ème restant s'***abstiennent** ***(a)***. *Chez les sportifs « loisir » **(l)**, **1/8**ème votent* **oui** ***(o)***, ***6/8**ème votent* **non** ***(n)*** *et **1/8**ème s'***abstiennent** ***(a)***.

- **a)** Donner la matrice (de transition) du premier canal T_1, contenant les probabilités conditionnelles $P(V_1 \mid S)$ du résultat du vote du premier tour V_1 (à valeurs dans {**o, n, a**}) sachant la catégorie S (à valeurs dans {***c, l***}) ;
- **b)** Calculer les probabilités associées à la variable V_1 et en déduire l'entropie $H(V_1)$;
- **c)** Calculer l'entropie conditionnelle $H(V_1 \mid S)$;
- **d)** Calculer l'information mutuelle $I_1 = I(S ; V_1)$.

2) <u>Deuxième tour de vote</u> (2,5 points) : *dans un second tour de vote, les sportifs qui ont voté oui (**o**) confirment leur vote ; de même pour les non (**n**) ; les abstensionnistes (**a**) se partagent <u>en égale proportion</u> entre oui (**o**) et non (**n**).*

- **a)** Donner la matrice du deuxième canal T_2, contenant les probabilités $P(V_2 \mid V_1)$ du résultat du vote du second tour V_2 (à valeurs dans {o, n}) sachant le résultat du premier tour V_1 ;
- **b)** Calculer l'entropie $H(V_2)$.
- **c)** Calculer l'entropie conditionnelle $H(V_2 \mid V_1)$;
- **d)** Calculer l'information mutuelle $I_2 = I(V_1 ; V_2)$.

3) <u>Globalement</u> (3 points):

- **a)** Donner la matrice de transition T_{12} contenant les probabilités $P(V_2 \mid S)$;
- **b)** Calculer l'entropie conditionnelle $H(V_2 \mid S)$;
- **c)** Calculer l'information mutuelle $I_{12} = I(S ; V_2)$.
- **d)** Interpréter la propagation de l'information de la catégorie de sportifs au travers des deux votes en comparant 2 informations mutuelles (que vous préciserez).
- **e)** **Pour une proportion p** de compétiteurs différente de 1/2, aurait-on pu avoir une information mutuelle I_{12} supérieure ? Justifier et interpréter en terme de capacité de canal.

<u>Exercice III :</u> Codage de Hamming (5 points)

*On considère un code linéaire bloc de Hamming, C, tel que **k = 11** et **n = 15**.*
On rappelle qu'un code de Hamming est un code <u>capable de corriger les erreurs simples</u>.

1) Indiquer : le **rendement** du code, les **dimensions** de la matrice génératrice **G** et de la matrice de parité **H** (appelée aussi matrice de contrôle), ainsi que le **nombre de mot-codes** (1,5 point) ?

2) Selon le 2ème théorème de Shannon, ce code pourrait-il être un candidat potentiel intéressant sur un canal binaire symétrique de capacité 0.5 bit (0.5 point) ?

On note **y** le mot-reçu en sortie d'un canal binaire symétrique.

3) Rappeler (en quelques mots) comment on calcule le **syndrome s (**à partir du mot-reçu **y)**, et à quoi il correspond lorsqu'il y a une erreur simple de transmission dans **y** (1 point).

4) En déduire comment on peut construire une matrice de parité pour ce code C, et donner l'exemple d'une matrice de parité **H** sous forme systématique (1 point).

5) Indiquer (sans la calculer) comment on peut obtenir la matrice génératrice **G** associée à votre matrice **H**, et donner le mot-code **c** associé au message **u** = [10000000000] (1 point).