Théorie de l'Information

1ère année

Examen de première session

POLYCOPIÉ, NOTES DE COURS ET CALCULATRICES AUTORISÉES
Le sujet est composé de 4 problèmes indépendants. Le barème est donné à titre indicatif

13/01/2021
9h-11h

CONSIGNES :

1. Écrire de préférence au stylo noir pour un meilleur contraste.
2. Recopier et signer l'engagement sur la première page de votre copie :
   Je soussigné (nom prénom) m'engage à effectuer seul mon évaluation, selon les modalités diffusées par l'enseignant. Je suis conscient que toute ressemblance avec une autre composition peut-être suspectée de plagiat et qu'agir en ingénieur responsable fait partie des compétences d'un ingénieur INP
3. Pendant l'épreuve vous devez être connecté à la réunion zoom " Théorie de l'information : Examen " . En cas de problème technique ou autre vous pourrez communiquer avec un enseignant via le chat.
   `https://grenoble-inp.zoom.us/j/95952327658`
   ID de réunion : 959 5232 7658
   Code secret : 366835
4. Noter sur chaque feuille votre nom, votre prénom et le numéro de page/ nombre total de page

5. à la fin de l'épreuve vous avez 15 minutes pour
   (a) Scanner votre copie sous format pdf (toutes les pages seront incluses dans un fichier unique, elles seront mises dans l'ordre et leur orientation doit permettre de lire directement votre copie sans avoir besoin de la modifier)
   (b) Enregistrer le fichier avec la nomenclature suivante :
   TI_Examen_numéro de groupe_nom_prénom_numéro étudiant
   exemple :**TI_Examen_groupe1_Dupond_Jean_41901501**
   (c) Déposer ensuite votre fichier sur Chamillo " Ensimag 3MMTINF Théorie de l'information " -> espace travaux ->
   (d) En cas de problème seulement envoyez votre fichier à l'enseignant encadrant votre groupe de TD en indiquant bien comme sujet :
   TI_Examen_numéro de groupe__nom_prénom .
   — groupe 1 et 2 : laurent.ros@grenoble-inp.fr
   — groupe 3 : olivier.michel@grenoble-inp.fr
   — groupe 4 : romain.couillet@gipsa-lab.grenoble-inp.fr
   — groupe 5 et 6 : michel.celette@grenoble-inp.fr
   — groupe 7 : alexandra.steinhilber@univ-grenoble-alpes.fr
   — groupe 8 : steeve.zozor@gipsa-lab.grenoble-inp.fr
6. En cas de manque de connexion réseau vous pourrez nous joindre à un de ces deux numéros
   — 04 56 52 96 35 (M . Celette)
   — 04 76 82 62 57 (M Ros)

Bon Travail à tous

1 CODAGE SOURCE (5.5 points)

Soit une source simple S d'alphabet $\mathcal{A}_S = \{a,b,c,d,e\}$ de 5 lettres, de distribution de probabilités Ps = $\{p(a);p(b);p(c);p(d);p(e)\}$.

Dans les 3 premières questions, la distribution de la source est :

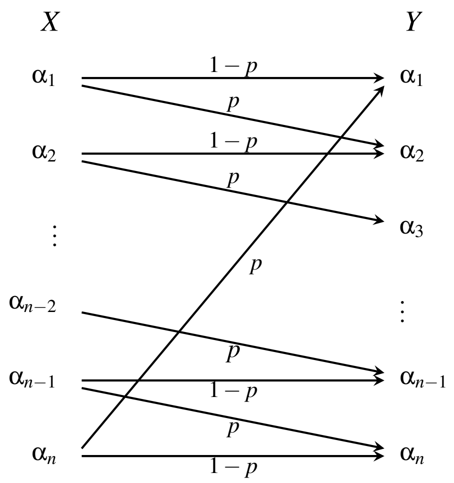
s	a	b	c	d	e
$P(S=s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{9}{16}$

1. Quelle est l'entropie, et la redondance de cette source S ?
2. Construire CH un code de source binaire de Huffman pour cette source.
   Quelle est l'efficacité de CH ? .
3. Construire CT un code de source ternaire instantané le plus efficace possible pour cette source. On prendra comme alphabet de codage $\mathcal{D} = \{0,1,2\}$
   Quelle est l'efficacité de CT ?
4. On suppose ici le code de source binaire C' = $\{1;10;100;1000;10000\}$
   (a) Ce code est-il déchiffrable (justifier) ?
   (b) Est-il possible de le remplacer par un code instantané ayant le même jeu de longueurs $\mathcal{L} = (l_1 = 1; l_2 = 2; l_3 = 3, l_4 = 4; l_5 = 5)$? (justifier)
   (c) Existe-t-il une distribution P'_S de la source permettant une efficacité du code C' de 100% ?
   - Si oui, donner une telle distribution, puis l'exemple d'un code C" efficace à 100%,
   - Si non, justifier, puis expliquer quelle(s) longueur(s) il faudrait modifier pour que ce soit le cas et donner alors la distribution P'_S.

2 CANAL TYPEWRITER (6.5 points)

Soit $n \geq 2$ et le canal suivant, appelé parfois canal typewritter.

En d'autres termes, pour $k = 1$ à n, le symbole α_k est transmis en le même symbole α_k avec une probabilité $1-p$ et en α_{k+1} avec une probabilité p, avec la convention $\alpha_{n+1} = \alpha_1$. On notera $\pi_k = P(X = \alpha_k)$, avec la même convention $\pi_{n+1} = \pi_1$. Pour $n = 26$ et les α_k lettres de l'alphabet, cela signifie que lors d'une saisie on tape sur la bonne touche du clavier avec probabilité $1-p$, et sur la suivante avec probabilité p ; les π_k sont les probabilités a priori des lettres du texte que l'on saisit.



1. Cas $n = 2$.
   De quel Type de Canal s'agit-il ?

2. Cas $n = 3$.

   (a) Donner la matrice de transition Π de ce canal.

   (b) Exprimer l'entropie conditionnelle $H(Y|X)$ en fonction de $h_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

   (c) Donner la loi de la sortie Y.

   (d) Quelle loi d'entrée maximise l'information mutuelle $I(X;Y)$?

   (e) Exprimer la capacité du canal en fonction de p.

3. Généraliser pour tout $n \geq 2$, les résultats des questions 2a), 2b) , 2d) et 2e). Pour la généralisation de la question 2d) on indiquera quelle propriété du canal permet d'y répondre sans avoir besoin de calculer la loi de la sortie.

4. Supposons que la matrice de transition du canal est de forme circulante (matrice carrée dans laquelle on passe d'une ligne à la suivante par permu-

tation circulaire)

$$\Pi = \begin{pmatrix} a_1 & a_2 & a_3 & \dots & \dots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \dots & \dots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_n & a_1 & \dots & \dots & a_{n-3} & a_{n-2} \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ a_2 & a_3 & a_4 & \dots & \dots & a_n & a1 \end{pmatrix}$$

$$où \;\; (\forall k \in [[1,n]])\, a_k \geq 0, \quad et \; \sum_{k=1}^{n} a_k = 1$$

On note $a = [a_1, a_2, \cdots, a_n] \quad et \quad H(a) = -\sum_{k=1}^{n} a_k log_2(a_k)$

(a) Exprimer la capacité de ce canal en fonction de n et $H(a)$

(b) Énoncer les propriétés du canal sur lesquelles repose la démonstration.

3 TRANSFORMATION DE L'ENTROPIE (3points)

Soit X une variable aléatoire discrète à valeurs dans $A = \{+1; +2; +4; +8\}$ d'entropie $H(X) = 2\ bits$

1. Quelle est la loi de probabilité de X
2. Calculer l'entropie de la variable aléatoire $Y = log_2(X)$
3. Montrer qu'il n'existe pas de fonction g telle que l'entropie de $g(X)$ soit strictement supérieure à $H(X)$
4. Déterminer une fonction f telle que l'entropie de $Z = f(X)$ soit non nulle et strictement inférieure à $H(X)$

4 CODAGE BLOC LINÉAIRE (5.5points) les deux parties peuvent être traitées indépendamment l'une de l'autre

1. On considère le code bloc linéaire de matrice génératrice sous forme systématique

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

Les mots de codes sont transmis au travers d'un canal binaire symétrique de probabilité de transition p.

(a) Quel est le rendement de code ?

(b) De combien de mots le code est-il composé ? Donner le liste des mots de code.

(c) Déterminer la capacité de détection et de correction d'erreur de ce code.

(d) Donner la matrice H de contrôle de parité.

(e) 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)

(f) En sortie de canal on observe la séquence $y = 1001001$

i. Déterminer le syndrome de y et la liste des séquences ayant le même syndrome

ii. Est-il possible de corriger la séquence y ? (si oui donner la correction proposée, si non justifier votre réponse)

iii. Commenter

2. **Bonus** On considère un code bloc linéaire de matrice génératrice sous forme systématique $G = \begin{pmatrix} I_k & P \end{pmatrix}$ où I_k est la matrice identité de rang k Sachant que $m_1 = 1000000$ et $m_2 = 1100000$ ont pour syndromes respectifs $\sigma(m_1) = 1010$ et $\sigma(m_2) = 0110$, que $m_3 = 1110000$ est un mot de code

Déterminer les syndromes de $m_4 = 0100000$ et $m_5 = 0001100$ puis justifier que $c = 0101100$ est un mot de code